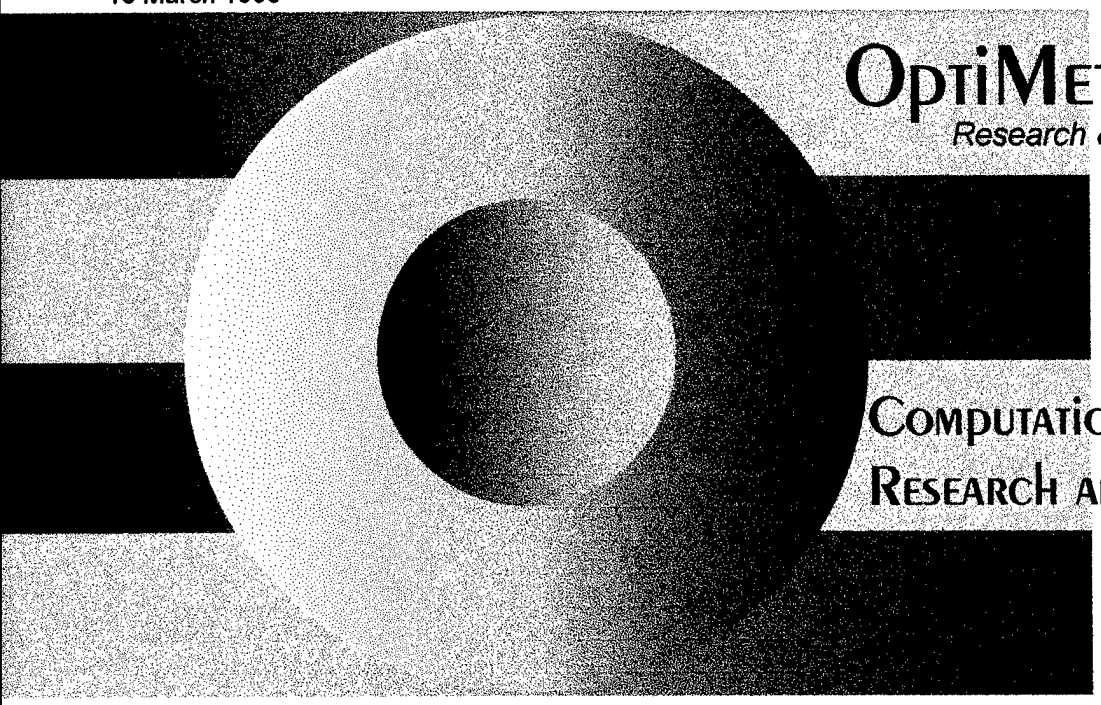
OMI-612

Phase II Final Report

13 March 1998



OptiMetrics, Inc.

Research & Engineering

Computational Vision Model Research and Development

Prepared For:

Commander
U.S. Army Tank-Automotive Command
ATTN: AMSTA-TR-S, Robert Karlsen
Warren, MI 48397-5000

Prepared By:

George H. Lindquist
Allyn W. Dunstan
J. Richard Freeling

Under Contract No. DAAE07-96-C-X053
CDRL Sequence No. A002
SBIR Report

OptiMetrics, Inc.
3115 Professional Drive
Ann Arbor, MI 48104-5131
http://www.omi.com

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	13 March 1998	Final Technical Report; 3/96 – 3/98

4. TITLE AND SUBTITLE
Computational Vision Model (CVM) Research and Development

5. FUNDING NUMBERS
DAAE07-96-C-X053

6. AUTHOR(S)
George H. Lindquist , J. Richard Freeling, Allyn W. Dunstan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
OptiMetrics, Inc.
3115 Professional Drive
Ann Arbor, MI 48104-5131

8. PERFORMING ORGANIZATION REPORT NUMBER
OMI-612

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Tank-Automotive Command
ATTN: AMSTA-TR-S, Robert Karlsen
Warren, MI 48397-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12A. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12B. DISTRIBUTION CODE
Distribution Statement: A

13. ABSTRACT (Maximum 200 words)

In the present effort, we have built a new and improved model for the prediction of target acquisition. It is based on the latest models of early vision processes in the human visual system. It is a multi-channel model, based on three color opponent channels, and two temporal channels. Each channel is further subdivided into a set of multi-resolution channels. Computations are based on contrast ratio. Local energy values for these dimensionless contrast images are computed, and normalized by the sum of eye noise and clutter. The detectability metric is then computed from these energies.

This model has been applied to the problem of detecting cars approaching intersections, and to the detection of mobile ground targets in imagery taken from an airborne first generation FLIR system. Reasonably good correlations were obtained.

Several alternative studies were also undertaken, along with a critique, in hindsight, of the ideas and philosophy used in the development of the model.

14. SUBJECT TERMS
Visual model
Detectability
Target/Background
Detection Modeling
Target Acquisition Model

15. NUMBER OF PAGES
58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

OMI-612; 13 March 1998

Final Report

For

Computational Vision Model Research and Development

Contract No. DAAE07-96-C-X053

CDRL Sequence No. A002

Prepared for:

Commander
U.S. Tank-Automotive Command
ATTN: AMSTA-TR-S, Robert Karlsen
Warren, MI 48397-5000

Prepared by:

George H. Lindquist
J. Richard Freeling
Allyn W. Dunstan
OptiMetrics, Inc.
3115 Professional Drive
Ann Arbor, MI 48104-5131

Table of Contents

Table of Figures

Appendices

1. Introduction

The purpose of this effort was to research, develop, and demonstrate an improved computational vision model to simulate human performance in image analysis tasks. The continued development and improvement of the National Automotive Center-Visual Performance Model (NAC-VPM) on this program represents the culmination of more than five years of biologically-inspired CVM-related research sponsored by the U.S. Army Tank-Automotive Research, Development and Engineering Center. OptiMetrics gratefully acknowledges the support and contributions of Drs. Grant R. Gerhart and Thomas Meitzler at TARDEC, and of Mr. Dennis Wend at the NAC.

1.1 Visual Target Acquisition

The NAC-VPM is a model of observer performance as defined in terms of the ability of a soldier to visually acquire a target, either by observing the target directly or by viewing the target on a display resulting from the collection of a scene by an electronic sensor. Visual performance is measured in terms of the degree to which the target object is distinguishable, in some military sense, from its surroundings. In military applications, target acquisition is described by several discrete tasks: Detection, Identification and Recognition, each of which have particular definitions in terms of the classes of objects between and within which distinction is made in the observation process.

These military measures of performance also have application in other areas. The best example, for the purposes of this effort, is in driving. The quantification of the ability of a driver to distinguish an oncoming vehicle in a fixed observational situation represents a problem which can be profitably addressed using military target acquisition models. This property of an object to be observed has been called its 'conspicuity'. Some other obvious 'conspicuity' applications are the visual prominence of:

- a particular object in a display ad.
- a warning sign or other attention-getting device.
- an object of a particular brightness and color against a particular natural background.

1.2 Prediction of Target Acquisition

The predictions of observer performance, dependent on the appearance of the target and its surrounding background, are important to the military for the purposes of surveillance, camouflage and targeting. These predictions are generally based on computational analysis of target/background images. However, because of the complexity and unknown nature of the

operations occurring within the human visual system and the inability to incorporate any cognitive information into these image-based calculations, such performance models have been only moderately successful taken over a full population of scenarios and situations and can be quite unsuccessful for particularly difficult targets.

1.3 Scope of the Present Modeling Effort

The NAC-VPM is based on the latest models of early vision processes in the human visual system. It is a multi-channel model based on three color opponent channels and two temporal channels. Each channel is further subdivided into a set of multi-resolution channels. Computations are based on contrast ratio images, one per channel, wherein the normalizing luminance is computed locally to each point in the image. Local energy values for these dimensionless contrast images are computed and normalized by the sum of a modeled value for the vision system's 'dark noise' and a locally modeled clutter value. The detectability metric is then computed as a weighted sum of these normalized channel energies. Reference 1 describes the NAC-VPM model in moderate detail. It inherits many of its details from its predecessor, TVM, and much the detail of NAC-VPM is described in the TVM Analyst's Manual (Reference 2).

This model has been applied to several problems. First, the model has been applied to the problem of detecting cars approaching intersections. Observer tests were run to measure the ability of an observer to detect vehicles at an intersection, as a function of distance, crossing speed and lighting conditions. NAC-VPM was able to predict the outcome of the observer tests to a correlation value of about 0.89 (Reference 1).

Second, the model was applied to the detection of mobile ground targets in imagery taken from an airborne first generation FLIR system. Good correlation between the VPM detectability/recognizability predictions and the human operator results were obtained. Correlation r-values in the high 70% range were achieved. These studies are reported in Reference 3. In the future, we hope to evaluate NAC-VPM's utility for modeling detectability of camouflaged targets in 3rd generation FLIR imagery.

Section 2 of this report contains an overview of image metrics for computational vision directed toward a more complete understanding of how the human visual system makes its detection decisions. In addition, two investigations were conducted into alternative methods of modeling target detection.

First, Section 3 describes an alternative view of imagery as three-dimensional solid objects and describes detection efforts in terms of differential geometries on the surface of these solids. This approach was applied to early measurements of the luminance contrast sensitivity of the

human eye. While the approach was successful, it appears that the conclusions to be made from this interpretation are no more insightful than those obtained by traditional methods.

Second, Section 4 describes an investigation of using extended Gabor filters as representations of the filtering done by the human visual system. While this direction seems to show some promise, the small number of applications we have made of these filters preclude any useful conclusions about them.

Finally, in Section 5, the philosophy behind NAC-VPM and TVM is revisited in hindsight, particularly in light of the successfulness of NAC-VPM. Based on this hindsight, we recommend in Section 6 some modifications to the approach taken to target detection modeling.

2. Receptive Field Organization and Its Implications for Visual Detection Metrics

2.1 Overview

Over the last seven to nine years an evolutionary step has occurred in the thinking concerning the human visual system's computational architecture. This has been driven predominantly by fundamental neurophysiological experiments on primates (References 4 and 5). The visual systems of the primate and human have proven to have a great number of similarities. Two major classes of retinal ganglion cells have concentrically organized receptive fields. Classically, the receptive field (RF) is defined as the area of visual space within which one can influence the discharge of a neuron. The RF is a central construct in the conceptual and analytical framework used by neuro-physiologists to study the function of visually responsive neurons, because it characterizes the transformation between the visual image and neuronal activity.

In recent years, the development of RF mapping techniques has facilitated characterization of RF's for neurons in the geniculo-cortical processing stream (Reference 6). Results obtained using this approach have resolved some long-standing questions concerning the origin of neuronal response properties, such as direction selectivity. These studies have revealed new aspects of RF structure posing new challenges for understanding and modeling the neural circuitry of the early visual pathways.

Beyond the low level vision pathways, into the inferior temporal cortex, a better understanding of the nature of the transformation and processing of low-level visual information for recognition decision making is now available. This understanding is illustrated in Figure 1. In the next section, we summarize some of these results and their impact on the developments within the present program.

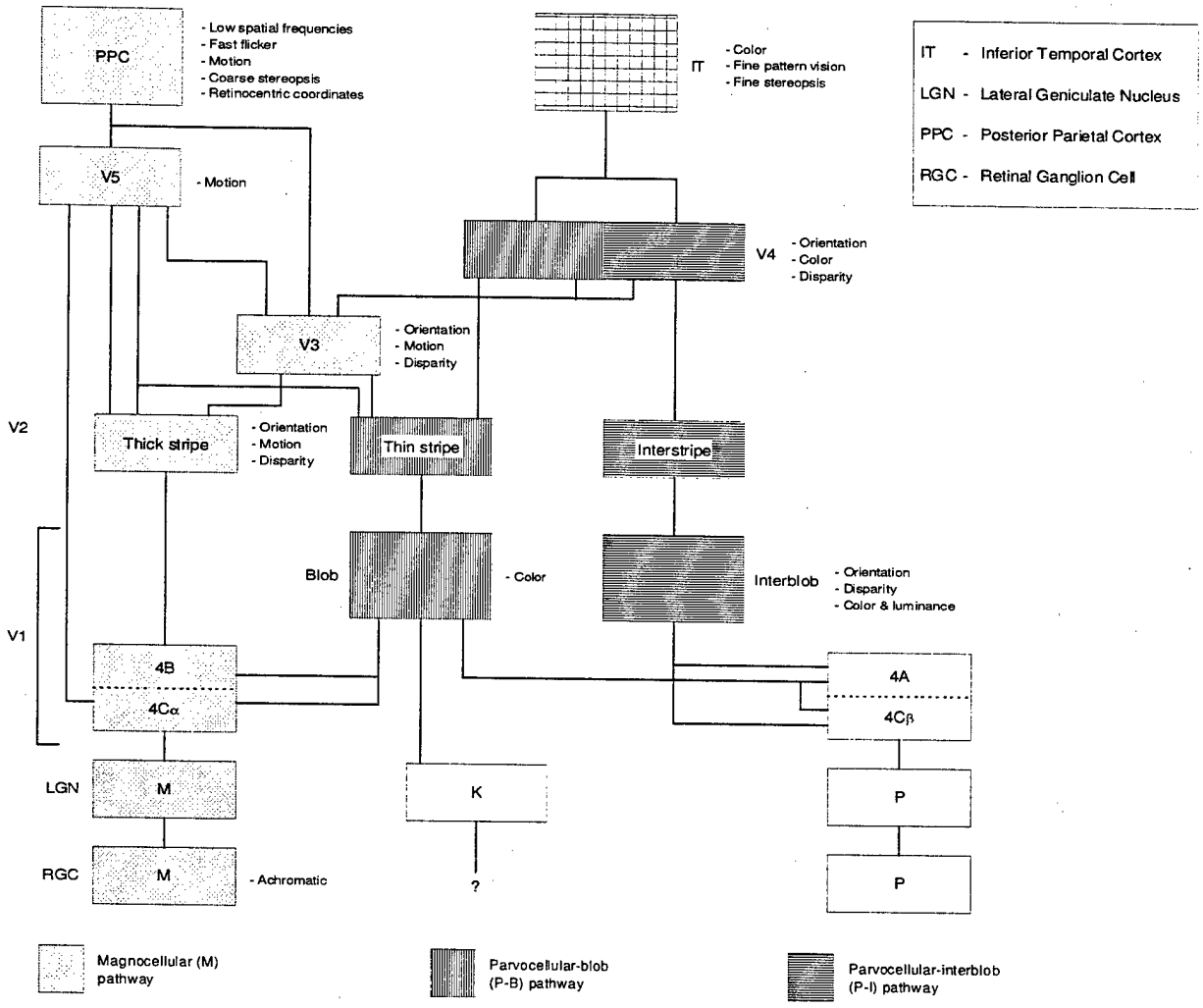


Figure 1. Sub-cortical and cortical pathways in the Macaque monkey (Reference 7).

2.2 The Lateral Geniculate Nucleus (LGN) System

Two primary classes of ganglion cells have concentrically organized receptive fields. These are widely known as the P- and the M-cells because of their different projections to the parvocellular (P) and magnocellular (M) laminae of the LGN. As the P and M cells penetrate the LGN they remain on distant, parallel pathways. Because the connection between the two classes of ganglion cell and their counterparts in the LGN is very well established, and the neurons in the LGN seem to have properties almost indistinguishable from those of the ganglion cell that drive them, the bulk of our understanding of early vision processing is based upon studies in the LGN.

The centers and surrounds of RF's of P-cells can be represented by Gaussian functions with a color-opponent organization superimposed on the spatial organization: center and

surround have different spectral sensitivities. There are two distinct subgroups of P-cells: those that have a 'red-green' color opponency and those that have a 'yellow-blue' chromatic opponency. Near the fovea RF's of P-cells are very small, and single cones probably drive centers. With small cell bodies, P-cells account for 80% or more of retinal ganglion cells. The upshot of this arrangement is that P-cells (particularly the red-green opponent cells) respond well to achromatic spatial contrast except at low spatial frequencies; when the spatial frequency is low they respond well to temporally modulated chromatic patterns.

M-cells have the same center-surround arrangement as P-cells, although the center of the receptive field has a diameter 2-3 times larger and the chromatic opponency is weak at best. Within the range of spatial frequencies to which both P- and M-cells respond, M-cells are much more sensitive to achromatic contrasts. This advantage is more pronounced at higher temporal frequencies. M-cells, with their larger cell bodies, account for approximately 10% of the ganglion cell population.

2.3 Organization in Cortical Signal Pathways

In the monkey (and presumably in humans), the principal projections of the P- and M-pathways are to area V1 of the visual cortex. Within V1, M-cells project from the LGN into layer 4Cβ while P-cells project into layer 4Cα.

In V2, there seems to be three separate visual maps (Reference 8). Within the thick stripes, there is a visual orientation map; within the thin stripes there is a color map; and within the interstripes a disparity map. Adjacent stripes are responsive to the same region of the visual field. So there are three interleaved visual maps in V2, each representing a different aspect of the visual stimulus. As shown in Figure 1 above, the M pathway projects from layer 4B of V1 to the thick stripes of V2. The pathway projects from the blobs of V1 to the thin stripes of V2 and from the interblob region to the interstripes of V2.

The representation of the visual field that projects into V1 is retinotopic, i.e., near neighbor relationships in the visual field are preserved in the sublayers of V1. Several stages later in the visual system, at the inferior temporal cortex (IT), the receptive fields are relatively independent of retinal location, and neurons can be activated by a specific stimulus, such as a face, over a wide range of retinal locations. So, along the visual system pathways, the pattern of excitation that reaches the eye must be transposed from a retinotopic coordinate system to a coordinate system centered on the object itself. At the same time that coordinates become object centered, the system becomes independent of the precise metric regarding the object itself within its own

coordinate system. Thus the visual system remains responsive to an object despite changes in its size, orientation, texture and completeness. Single-cell studies in monkeys suggest that for faces processing these transformations occur in the IT (see Figure 1 above).

2.4 Global Effect of Local Contrast

Brightness perception in complex, stationary scenes has played a key role in development of our hypothesis concerning a new, somewhat geometric, approach to the human visual system's computational architecture. Before describing this approach (Section 3, below), a brief overview of what is currently accepted concerning the influence of local contrasts on global or area brightness perception is provided.

There are several factors which, together, help to determine how humans perceive brightness of an enclosed area.

(I) For an enclosed area, brightness perception is determined primarily by the average local contrast at the boundary or edge of the area.

(II) The boundary between two regions of differing contrast influences their brightness perception, but the strength of the boundary's contribution decreases with increasing distance.

(III) Local luminance modulation, which simulates an edge, can generate area contrast.

(IV) Small luminance gradients from one spatial region to another are not particularly effective in determining local brightness.

These gradual changes are mostly filtered out by the visual system. It is still an open question as to how local contrast at edges may help produce brightness contrast over large stimulus areas. To date, no neurophysiological correlate of area contrast, either in the retina or in the LGN, has been identified.

It has been observed in cortical receptive field studies that individual neurons provide no simple answer, since neurons whose receptive fields fall within a homogenous area are not active at all (Reference 9). Thus, one could argue that area contrast is produced because oriented neurons code the direction of contrast gradients at each border and that this information is somehow propagated from one border to another across space. Thus, the brightness of an area of homogenous luminance would be extrapolated from the directions of the contrast gradients at the border.

But where is neuronal activity bound to perception of light and dark? If boundaries (orientation and high frequency channels) are key, the P-cellular pathway must play a significant role providing the needed information to an area contrast perceptual mechanism (see Figure 1

above). However the P-pathway does not provide the high contrast sensitivity associated with strong edges or boundaries. This high contrast sensitivity, propagated via the M-pathway, is merged with P-pathway signals only in the upper layers of V1 and within V2.

Thus any perception of brightness and darkness may be carried out in multiple layers of the visual cortex (at least layers V1 and V2) by processes which become more and more selective and which involve fewer and fewer (but increasingly selective) neurons. And hence the perception of area brightness is likely taking place well along the cortical pathways, perhaps into V4 and the inferior temporal cortex where higher level cognitive processes (recognition) are thought to operate.

In Section 3 below, a computational methodology is presented which incorporates the transition to more object centered information, while retaining certain retinotopic data; at the same time, a synthesis is being performed to yield a geometric object which appears to be fundamental in higher level discrimination. As evidence of the role that fundamental geometry may play in perception (higher-level discrimination tasks), the proposed methodology is utilized to provide a prediction of Blackwell's historic perception test results (Reference 10).

3. Riemannian Manifolds and Visual Perception

3.1 Background

The machine vision boom of the 1980's stimulated substantial efforts directed at understanding two and three-dimensional images from a purely geometric perspective. Shape from shading and image invariance are two examples of fundamental work born out of this era which have provided insights into a promising new direction for understanding visual perception in humans. In the following, discussions of shape from shading and the concept of image invariants are provided. Briefly mentioned are insights gleaned from these individual approaches to machine understanding of images, which may have potential to enhance our understanding of human visual perception. This section concludes with a description of a promising approach to human visual perception that integrates these insights, based upon two-dimensional, intrinsic, surface geometry. Extension from static to dynamic imagery is natural and will only be briefly described.

3.2 Shape from Shading

Since brightness variation is an important component of the information our visual system utilizes for decision making, a portion of the present effort was devoted to understanding the shape from shading approach for deriving surface data from brightness variations.

Shape from shading brings together several simplifying assumptions concerning the relation of radiation incident on a surface to the reflectance map of that surface. Very simply, what is captured by shape from shading analysis is surface normal behavior from brightness changes over the surface. Additionally, under certain conditions, surface height can also be determined. Shape from shading has proven to be an insightful means of exploiting these brightness variations for deriving surface contour information. A very good description of this specialized area, its assumptions and some applications, is given in Horn (Reference 11). Fundamental to this approach is the relationship

$$L(x,y) = R(p,q) \cdot E(x,y) \tag{1}$$

of image radiance, L(x,y), to the reflectance map, R(p,q), through the irradiance (illumination), E(x,y), where the gradient of the surface z(x,y) at the point (x,y) is (p,q):

$$p(x,y) = \frac{\partial z(x,y)}{\partial x} \tag{2}$$

$$q(x,y) = \frac{\partial z(x,y)}{\partial y} \tag{3}$$

The basic dependence of surface brightness upon surface orientation is contained in solutions to Equation 1, a two-dimensional, first order nonlinear partial differential equation. Shape from shading exploits these variations in brightness (shading) to obtain estimates of surface orientation. With estimates of p(x,y) and q(x,y) for each pixel in an image, the surface height above some reference point $z(x_0, y_0)$ can be obtained from the path integral

$$z(x,y) = z(x_o,y_o) + \int_{x_o,y_o}^{x,y} (pdx + qdy) \tag{4}$$

In this expression for surface height, closed paths may be used which would allow conversion of the path integral, via Stokes Theorem, to a surface integral. This surface integral vanishes for integrable surfaces (∂_y p(x,y) = ∂_x q(x,y)), i.e., this integral around any path is independent of the choice of the path. However, this behavior is a coordinate-dependent result. Because the human visual system is able to estimate rather complicated (integrable) shapes, i.e., facial contours, from a single picture, photograph, etc., it would seem shape from shading would be of very limited utility in the perception of surfaces by our visual system. On the other hand, the most basic information available to the visual system consists of these brightness variations in every image it encounters.

In an attempt to augment shape from shading and perhaps gain some insight into the robust behavior of our visual system, our effort focused on image information as purely geometric data. With no assumptions about the behavior of the surface reflectance, this would allow the visual system to thus be 'fooled' by manipulation of surface geometry and surface reflectance characteristics. While this is clearly undesirable from a machine vision perspective (reverse engineering of brightness data to obtain surface shape), it appears to be a reasonable visual system performance model which is compatible with the current understanding of primate visual system processing. As a coordinate independent description, it relies on elementary differential geometry and thus preserves local (brightness) metric relationships while allowing for determination of large area contrasts. This geometric approach to visual system performance is described below, followed by a comparison of predicted detection probability curves for uniform disks with Blackwell's performance data.

3.3 Geometry of Brightness Images

It is envisioned in the proposed geometric approach to visual perception that following the

early visual system processing, each output image plane, labeled by temporal, color opponent, orientation, and spatial frequency indices, would be treated as a two dimensional surface in three dimensional space,

$$z = f(x,y), \tag{5}$$

with indices omitted. The usual Euclidean line element, in Cartesian coordinates,

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{6}$$

induces a (non-Euclidean) metric in the surface, f(x,y), which, with a trivial change of variables $(x,y) \leftarrow (u_1,u_2)$, yields the line element in the surface as

$$ds^2 = g_{ij}\, du_i\, du_j \tag{7}$$

with, repeated indices summed, and

$$g_{ij} = \begin{bmatrix} 1+f_1^2 & f_1 f_2 \\ f_1 f_2 & 1+f_1^2 \end{bmatrix}_{ij}, \tag{8}$$

where

$$f_i = \partial_i f(u_1,u_2).$$

The right hand side of Equation 7, above, is the first fundamental quadratic form of the surface (Equation 5) with which the arc length of any curve embedded in $f(u_1,u_2)$ may be determined. If the vision system were dependent solely on the information in g_{ij}, surfaces where the brightness map yields the same g_{ij} would be ambiguous. For example, surfaces which could not be distinguished are those that can be cut and placed flat without stretching, compressing or tearing, e.g., cylinders, cones, etc.

To eliminate this ambiguity, a second fundamental quadratic form is introduced, d_{ij} (Reference 12). Utilizing these image-based geometric quantities, it is hypothesized that the human visual system is able to make detection/recognition decisions by comparing components of a vector over regions of an image surface for which the Riemann curvature tensor, R_{1212}, is non-zero. This comparison process is known as parallel transport. L. D. Landau and E. M. Lipshitz (Reference 13) describe this process, along with the definition of R_{1212}. Comparing a vector to itself at points over an image surface measures the change in direction of the vector due to shaping of the image surface. This change is given by

$$\Delta A_k = \int_C \Gamma^i_{kl} A_i du^l \tag{9}$$

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

where C is any closed contour in the image surface, Γ^i_{kl} are the Christoffel symbols of the second kind constructed from the g_{ij}; and A_i , i= 1,2, are the components of a two dimensional vector, $\underline{A}(u_1,u_2)$ whose orientation change, ΔA_k , is hypothesized to correlate with discrimination performance. Any region of the image plane f, with non-zero curvature ($R_{1212} \neq 0$), induces a change in the orientation of the vector $\underline{A}$ when $\underline{A}$ is parallel transported around the path C enclosing the region. Figure 2 shows the input-output data flow for Equation 9. In this figure, the vector $\underline{A}$ is fed to the mid-level visual processor from a high-level, long-term visual data archive. The mid-level processor receives input of regions of interest and their associated boundaries along with the Christoffel symbols for the designated image plane. With the high- and low-level inputs, the visual processor provides the necessary multiplications and adds to yield the ΔA_k. The contour integral in Equation 9 and Figure 2 is probably implemented, via Stokes' Theorem, as a surface integral. In this case, mid-level processing amounts to the components of $\underline{A}$, weighted by the curvature tensor, summed over the area enclosed by curve C. That this is reasonable will be seen below in the application of Equation 9 to Blackwell's performance data. There it will be shown that Equation 9 reduces to Ricco's Law.

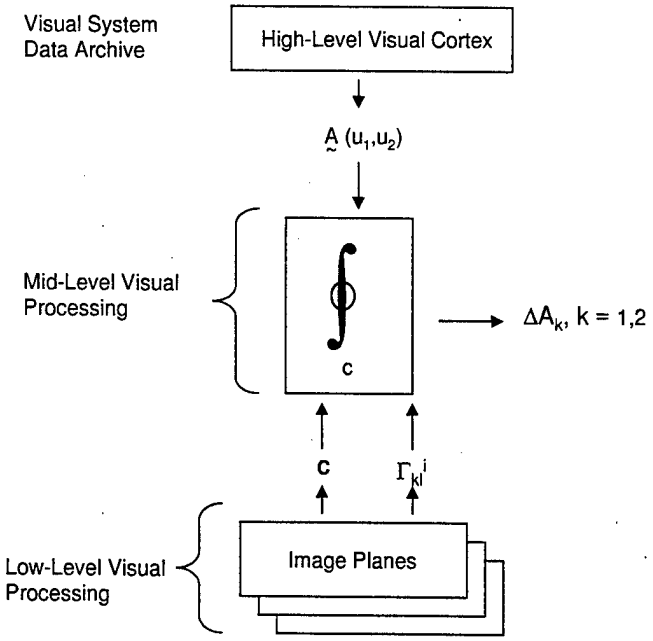


Figure 2. Input-output data flow involved in computation of image curvature induced orientation changes.

Note the similarity of Equations 4 and 9; they both have the form

$$\int_C \left\{P(u_1,u_2)du_1 + Q(u_1,u_2)du_2\right\} \qquad (10)$$

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

which is a path integral over a curve C. By relying on the intrinsic geometry of an image regarded as a two dimensional geometric surface embedded in a three dimensional space, Equation 9 is non-zero independent of the integrability condition ((∂_y p(x,y) = ∂_x q(x,y)) .

3.4 Application to 1946 Blackwell Perception Tests

In 1946 Blackwell published a landmark paper in human perception testing (Reference 10). In these experiments, a spot (disk) of light was projected onto a white screen located approximately sixty feet from a group of observers who reported whether the spot had been seen. Spots of varying size were used against various background brightness levels. In this section, an analytic result for comparison with Blackwell's contrast threshold versus spot size will be presented for varying levels of background brightness.

To apply the results of Section 3.3, a synthetic image of a disk in a uniform background was generated by convolving a radially symmetric step function with a Gaussian of width t. The result, in polar coordinates (u_1,u_2), is very closely approximated by the following expression,

$$z(u_1,u_2) = \frac{h}{2\sqrt{\pi t}} \int_{-\infty}^{\rho} \exp\left(-\frac{(u_1 - \eta)^2}{4t} \right) d\eta \qquad (11)$$

where, as shown in Figure 3, h is the disk height, ρ its radius and u_1 is the radial coordinate. With this as the image, the required Christoffel symbols are

$$\Gamma_{22}^1 = -\frac{u_1}{1+z'^2}, \Gamma_{11}^1 = -\frac{z'z''}{1+z'^2} \qquad (12)$$

where the superscript prime denotes a derivative with respect to u_1. For the vector $\underline{A}$ (Equation 9), a simple choice is made: a constant vector in the u_1 direction of magnitude A_0 is used:

$$\underline{A}\,(u_1, u_2) = \{A_0, 0\}. \qquad (13)$$

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

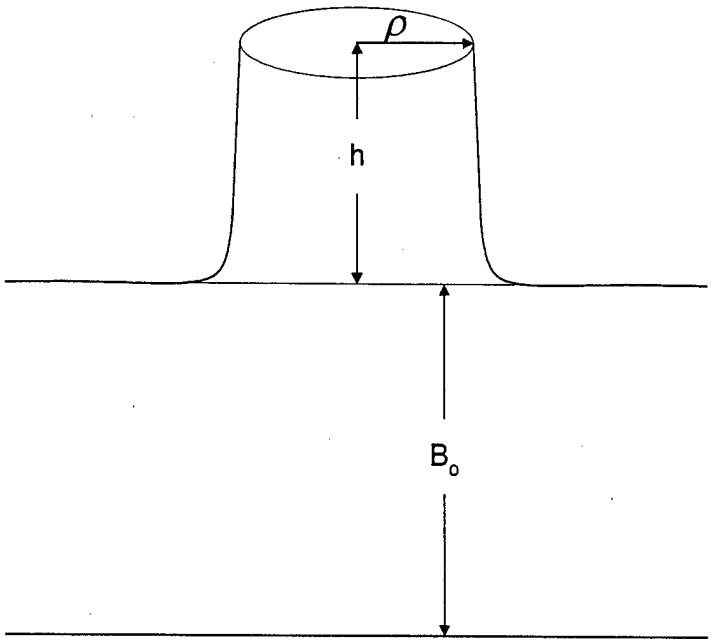


Figure 3. Smoothed disk of radius ρ and contrast h in uniform background, B_o.

The integration path, C, in Equation 6 was taken to be a circular path of radius equal to the disk radius and centered on the disk. With these choices, the change in orientation of the constant vector A upon parallel transport around the disk perimeter was found to be

$$\Delta A_1 = 0,$$

$$\Delta A_2 = \frac{\pi}{2}\left(\frac{h}{B_o}\right)^2 A_o r^2 \frac{\left(\frac{h}{B_o}\right)^2 + 8\pi t}{2\left(\left(\frac{h}{B_o}\right)^2 + 4\pi t\right)^2} \qquad (14)$$

up to an overall scale factor. In this last expression, h has been scaled by the background brightness for ease of comparison with Blackwell's Figure 16. In Figure 4, level curves of the magnitude of ΔA_2 are plotted as a function of $\log(\rho/B_0)$ vs. $\log(h/B_0)$. In this plot, A_0 was taken to be the square root of h and t has the value 0.004. As in Blackwell's plot, approximately three orders of magnitude in disk size and more than six orders of magnitude in contrast are represented in Figure 4. The level curves represent contours of constant B_0. Figure 4 could be directly compared to Blackwell's Figure 16, if it were calibrated against contrast threshold predictions at 50% probability. This would seem to be an interesting task for future effort.

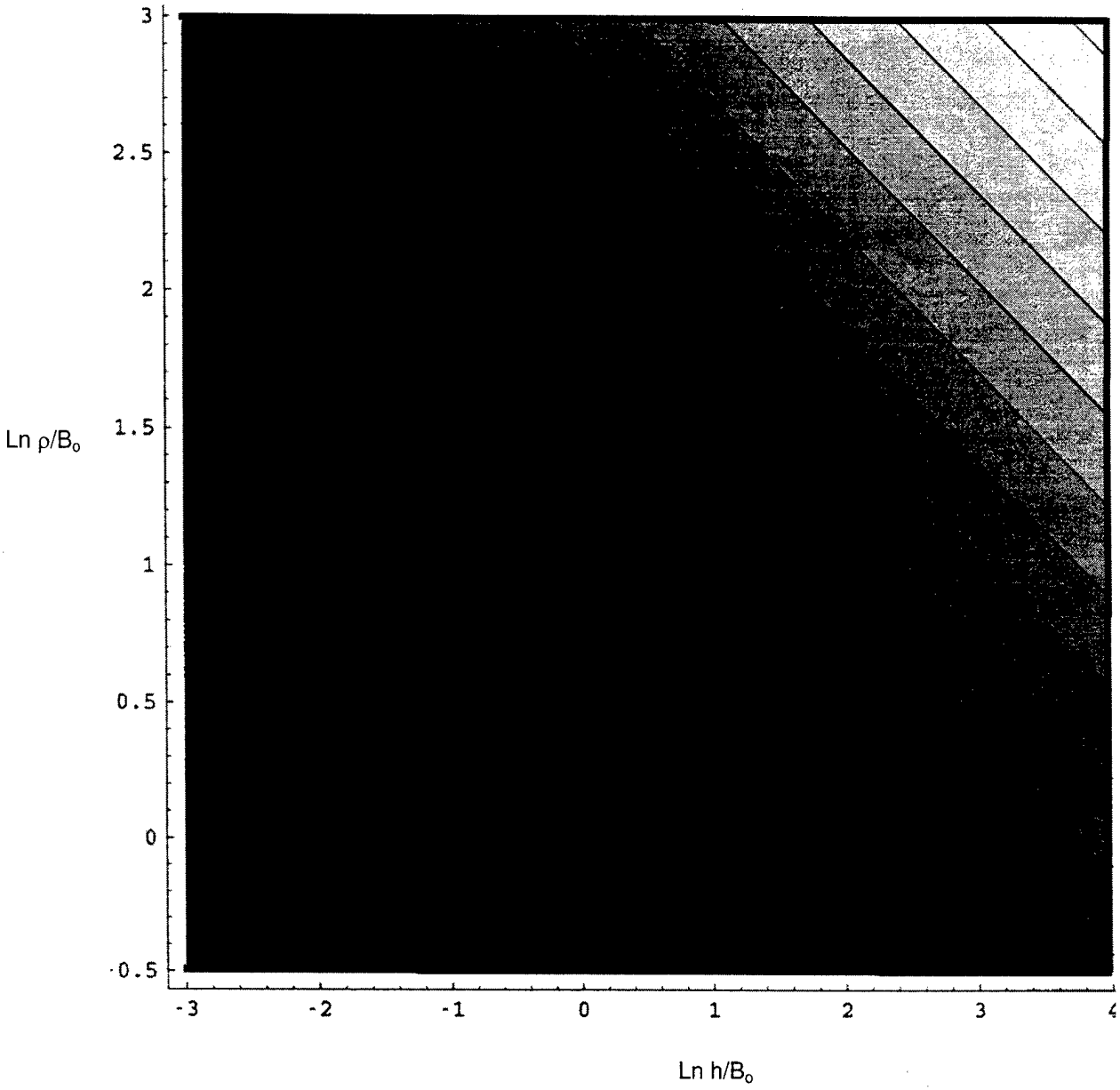


Figure 4. Level curves of $|\Delta A_2|$ plotted as a function of ln (ρ/B_o) vs. ln (h/B_o). For this plot, t = 0.004, $A_o = \sqrt{h}$.

There are several interesting features worth noting concerning the prediction above. The linear portion of each level curve represents the fact that the product of area and brightness is a constant. Any stimulus on the linear portion of the curve is effectively a 'point source'. The point at which a level curve departs from linearity corresponds, according to Blackwell, to a fundamental property of the human eye. In this vein it is interesting that, from the expression above, as the degree of blurring, governed by the parameter t (Gaussian width), decreases, the onset of upward curvature shifts to left in Figure 4. This appears to be at least consistent with Blackwell's

observation that the onset of curvature in Figure 4 is a characteristic of the human eye. To the extent that greater blurring can be ascribed to the aging eye, one could predict the decrease in the linear portion of Figure 4 for the aging eye by increasing the Gaussian width, t.

Thus this geometric approach to imaging does appear to shed light on the form of the contrast sensitivity of the human vision system.

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

4. Extended Gabor Filters and a True Metric

4.1 Gabor Filters as Harmonic Oscillator Functions

In an attempt to identify meaningful new approaches to visual discrimination, the basic filtering process of the low-level visual system was examined with the idea of attempting to quantify its information content. This is, in effect, what the current VPM metric attempts to accomplish through a weighted pooling of filter channel outputs. As a starting point for this effort, Gabor's original work on communication theory was reviewed (Reference 14). In this pioneering paper, Gabor points out the utility, for his analysis, of the products of complex exponentials and Gaussians which have come to carry his name, 'Gabor filters'. He also points out that these functions are members of a larger class of functions, i.e., harmonic oscillator (HO) wave functions. HO wavefunctions are solutions to the one dimensional, linear harmonic oscillator Schrödinger equation and are of the form

$$\Psi_{n,\sigma}(z) = N(n,\sigma)\exp\left[-\frac{z^2}{2\sigma^2}\right]H_n\left(\frac{z}{\sigma}\right) \tag{15}$$

where H_n (z/σ) are Hermite polynomials of degree n and $N(n,\sigma)$ is a parameter dependent normalization constant. These functions are normalized, orthogonal and make up the (separable) solution to the two dimensional HO equation. These two dimensional solutions are of the form

$$\Psi_{n_1,\sigma_1,n_2,\sigma_2}(x,y) = \Psi_{n_1,\sigma_1}(x)\Psi_{n_2,\sigma_2}(y) \tag{16}$$

Berry pointed out in 1984 that there is an additional degree of freedom in Schrödinger equations which introduces a fundamental change in Equations 15 and 16 above (Reference 15). What Berry points out is that as the parameters of a Schrödinger equation change, its solutions acquire what he termed a geometric phase. This phase has been referred to in the literature as 'Berry Phase' (Reference 16). It is of the form

$$e^{i\alpha(x,y)} \tag{17}$$

where the functional, $\alpha(x,y)$, is given by

$$\alpha(x,y) = \int_{\gamma} \underline{B}(\underline{\Omega}) \cdot d\underline{\Omega}\,, \tag{18}$$

with the four dimensional parameter vector $\underline{\Omega}$ having the components

$$\underline{\Omega} = \{n_1, \sigma_1, n_2, \sigma_2\}. \tag{19}$$

The vector quantity, $\underline{B}$, termed the Berry vector potential or Berry connection is given as

$$\underline{B} \equiv i\Psi^*_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \cdot \nabla_{\underline{\Omega}} \cdot \Psi_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \tag{20}$$

where the gradient in this last equation is four-dimensional and is taken with respect to the parameters. The path of integration, γ, in Equation 18 is in parameter space and may be a closed path. In the case of closed paths, Equation 18 may be converted to a surface integral for evaluation which is a convenience in many instances. With Berry phase, Equation 16 would acquire the multiplicative phase factor in Equation 17, creating an 'extended' two-dimensional Gabor filter. Thus taking into account Berry phase, one may create 'extended' Gabor Filters which bring additional parameters, n_j (j=1,2), one for each of the two spatial dimensions and add, in a highly non-linear fashion, the geometric characteristics of the parameter space to the visual filtering process. This filter would have the form

$$\Psi_{n_1,\sigma_1,n_2,\sigma_2}(x,y) = \Psi_{n_1,\sigma_1}(x)\Psi_{n_2,\sigma_2}(y) \cdot e^{i\alpha(x,y)}. \tag{21}$$

Use of 'extended' Gabor filters offers a new and potentially fruitful path for further investigations of the computational architecture of low level visual processing, but falls short in providing insight into meaningful new metrics. Following Berry's fundamental paper, there have been new investigations into the broader geometric aspects of Berry's discovery which do provide a metric, a true metric, applicable to the parameter space of 'extended' Gabor filters. Page describes several new geometric structures and their relationship to Berry's original arguments (Reference 17). Included in this discussion is a metric, the Fubini-Study metric, which acts on the parameter space investigated by Berry. Applied to the four dimensional parameter space of 'extended' Gabor filters, this metric is a symmetric, second rank tensor, $g_{i,j}$, with the form

$$\begin{aligned} g_{i,j}(\underline{W}) = &\iint dx \cdot dy \cdot \partial_i \Psi^*_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \cdot \partial_j \Psi_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \\ &- \iint dx \cdot dy \cdot \partial_i \Psi^*_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \cdot \Psi_{n_1,\sigma_1,n_2,\sigma_2}(x,y) \iint dx' \cdot dy' \cdot \Psi^*_{n_1,\sigma_1,n_2,\sigma_2}(x',y') \cdot \partial_j \Psi_{n_1,\sigma_1,n_2,\sigma_2}(x',y') \end{aligned} \tag{22}$$

where i and j index the components of $\underline{\Omega}$. From its structure, $g_{i,j}$ can never give a negative

distance: it is a positive semidefinite metric. With this metric, the natural distance between any two points $\underline{\Omega}_A$ and $\underline{\Omega}_B$ in parameter space is given by

$$s_{AB} = \int_{\Omega_A}^{\Omega_B} \sqrt{g_{i,j}(\underline{\Omega}) \cdot d\underline{\Omega}_i \cdot d\underline{\Omega}_j} \, . \tag{23}$$

While this direction of investigation appears to hold promise, any further investigation must be delayed to a follow-on program.

5. NAC-VPM Revisited -- Potential Improvements in Linear Modeling of Target Detection

Despite our best efforts, the performance of NAC-VPM and its predecessors does not appear robust over the wide range of targets, backgrounds, and scenarios encountered in the real world. Though the model's performance can be tuned or calibrated to statistically mimic human detection performance for a limited and controlled set of target/background/viewing conditions, the calibration seems scene-dependent. This lack of robustness is shared by all computational vision models of which we are aware. To better understand the source of the problem and to suggest a better way forward in the future, we have analyzed the assumptions on which NAC-VPM is based. The results of that analysis presented here suggest certain changes in the overall approach to this problem; we believe that implementing these changes will lead in turn to improved model performance.

5.1 Domains of vision

The eye is basically an imaging sensor. Sensible inputs to the retina consist of image spectral radiance values, integrated over the solid angle subtended by the eye's pupil and over the spectral responses of the sensing elements present in the retina. Sensible inputs are collected as a function of direction, over the angular field observed by the eye. The spatial resolution of the eye varies greatly with position in the field and with illumination level. The eye's greatest contrast sensitivity occurs at a spatial frequency of 3 to 5 cycles/degree, but it possesses significant contrast sensitivity out to spatial frequencies as high as 70 cycles/degree. This would indicate that the eye has some ability to resolve objects as small as 1/140 of a degree, although transfer function measurements indicate that the eye mostly integrates signals for sources smaller than 0.1 degree. Similarly, the eye is capable of distinguishing signals as a function of time. Depending on brightness, the eye integrates visual stimuli over about 0.1 second, and is capable of resolving temporal signals longer than about 0.1 second.

It is known that the human visual system is capable only of observing colors in terms of the rules of mixing of three so-called primary colors. This both indicates, and is a result of the fact that the human eye has sensors having only three different spectral sensitivities. Thus, although the image viewed by the human eye can contain spectral content of arbitrary complexity, since its sensors have only three different spectral responses, all perception of color must be representable in terms of three values. The rules of color mixing can be cast in terms of a number of different sets of three spectral sensitivities. The most commonly used, primarily for convenience in making colorimetric calculations, are called the tri-stimulus functions, called X, Y

and Z. Any brightness value expressed in this, or any other equivalent set of tri-stimulus functions encompasses all of the ability of the eye to perceive brightness and color. Thus the human eye can be said to be sensitive to three colors or integrations over the visual spectral region, extending from about 430 to 700 nm.

Thus, any continuous encoding of the tri-stimulus values over the full visual field, capable of representing spatial variations out to something less than 70 cycles/degree, and temporal events longer than 0.1 second should contain all information which the eye is capable of sensing. These encodings serve as sufficient input to any visual performance model. For this reason, tri-stimulus images (in the form of RGB images) at appropriate spatial and temporal resolution should serve as adequate input to any visual performance model.

5.2 Historical attempts to model detection

Models for the sensitivity of the eye, based on visual perception experiments, are indeed visual performance models, in that they describe the threshold of visibility for certain idealized target shapes. Blackwell's classic data measured the ability of observers to detect discs of different sizes, as a function of both the size of the disc and its luminance difference with its surrounding (Reference 10). More recent studies have developed a whole model for the contrast sensitivity of the human visual system in both spatial and temporal domains (Reference 9). While such a model yields a beginning understanding of the sensitivity of the eye, it is not useful for detection prediction because it applies only for idealized uniform targets having some fixed contrast with the background (such as uniform discs or grating patterns). The problem of military importance involves complex targets of varying luminance and color, and which may be matched to their surroundings in either in luminance, color or texture or a mixture of the three. Such targets have no simply defined average contrast or size to which a contrast sensitivity model can be applied.

More complex models have been developed based on more complex, but still predefined targets (Reference 18). The four-bar pattern of Ratches and others has been heavily used to model target acquisition through displayed images from an auxiliary imaging sensor such as a FLIR or image intensifier (Reference 18). These models blend linear system theory (to treat the sensor processes) with linear system models of human vision (to treat the observation of the displayed images). A pattern of four equidistant parallel bars of 7 to 1 aspect ratio, separated by spaces equal to the bar width, results in a square target region which has been used as a surrogate for complex military targets. A large amount of experimental and analytical effort has been expended to relate an observer's ability to detect this bar pattern to the observer's ability to

perform detection and recognition tasks against complex target/background scenes. However this target pattern still requires the existence of an overall contrast between the target and its background. This requirement for a contrast makes the bar target unusable for the modeling of visual target detection, since many visual targets are detected merely by their texture contrast with their background.

5.3 Computational Vision Approaches

Computational vision approaches to the modeling of target acquisition generally attempt to simulate the chain of processes leading to visual understanding, from the collection of light, through sensing it, and continuing through all those processes which we feel certain are part of the visual process. The simulation is stopped at that point in the vision system beyond which precise operation is not known. The output(s) resulting from this 'knowable' simulation are then cast in terms of a signal power vector, and a corresponding noise process is associated with each element of the vector. The resulting vector of power signal-to-noise ratios then becomes the metric against which actual detection experiments are correlated. The result is a relationship between probabilities of making correct decisions and this metric vector.

While much is made of the non-linearity of the human visual system, most of the low-level characteristics of the eye demonstrate a linear system behavior in dimensionless contrast, or contrast ratio. Computational vision approaches to vision modeling build on these linear relationships.

For this discussion, we can define simple luminance contrast as simply the difference between two luminance values, a local luminance and a local average luminance. It thus has the units of luminance. Thus,

$$\text{Luminance Contrast} = \text{Local Luminance} - \text{Local Average Luminance} \tag{24}$$

The contrast ratio is the ratio between the luminance contrast and a normalizing luminance. Thus,

$$\text{Contrast Ratio} = \frac{\text{Luminance Contrast}}{\text{Normalizing Luminance}} = \frac{\text{Local Luminance} - \text{Local Average Luminance}}{\text{Normalizing Luminance}} \tag{25}$$

If the target is very small, the normalizing luminance should be the average luminance of the background. In general, the normalizing luminance appropriate to a particular point in the scene will also be some local average of the luminance about that point, and indeed may be identical to

the Local Average Luminance used to compute a luminance contrast. The maximum value of contrast ratio is determined by the maximum neural firing rate of neural pathways in the human visual system. While TVM, NAC-VPM's predecessor, used contrast ratio only indirectly, the current version of NAC-VPM is built solely on contrast ratio images.

As an example of the linear behavior of the eye, the fact that the eye's luminance sensitivity can be modeled in terms of a contrast threshold function, with dependence on luminance level, color and field size, is itself an argument for human vision as a linear process. The description of contrast sensitivity in terms of a minimum perceptible value for the dimensionless contrast, valid for a range of luminance values, implies linear behavior in dimensionless contrast, at least over that range of luminance values, and at least for small contrast changes. As an even more convincing example, the generally accepted ability to describe the eye's threshold luminance contrast as a reproducible function of either target disc size or gray scale spatial frequency implies linear behavior. The eye's contrast sensitivity can be adequately described in terms of a linear imaging sensor of contrast ratio, having spatial resolution described by a spatial impulse response, with a finite 'minimum detectable contrast' value, and with a maximum field size. The presence of a spatial impulse response implies spatial integration over objects smaller than the width of that impulse response, and the specification of contrast ratio sensitivity as a function of spatial frequency (a contrast ratio MTF).

Blackwell's original liminal contrast data as shown in Figure 5 can be explained in terms of a constant liminal contrast ratio for large stimuli (see Figure 5a) and in terms of a contrast ratio-area product for small size visual stimuli (see Figure 5b), at any luminance adaptation level. In Figure 5, the liminal contrast for small size stimuli is proportional to product of contrast ratio and size, (demonstrated by a log-log slope of –2) while for large size stimuli, the liminal contrast ratio approaches a constant value as the size of the stimulus increases. This indicates that, for the full range of Blackwell's measurements, the eye can be represented as a linear sensor in contrast ratio. Much of the earlier literature has attempted to rationalize these linear representations in contrast ratio with the photon detection processes and subsequent luminance adaptation processes occurring in the human visual system (References 19 and 20). However, for purposes of modeling target detection, the details of how contrast sensitivity is achieved by the human visual system is irrelevant. We need only build a linear model based on contrast ratio, and apply it to contrast ratio images. NAC-VPM is an example of such a model.

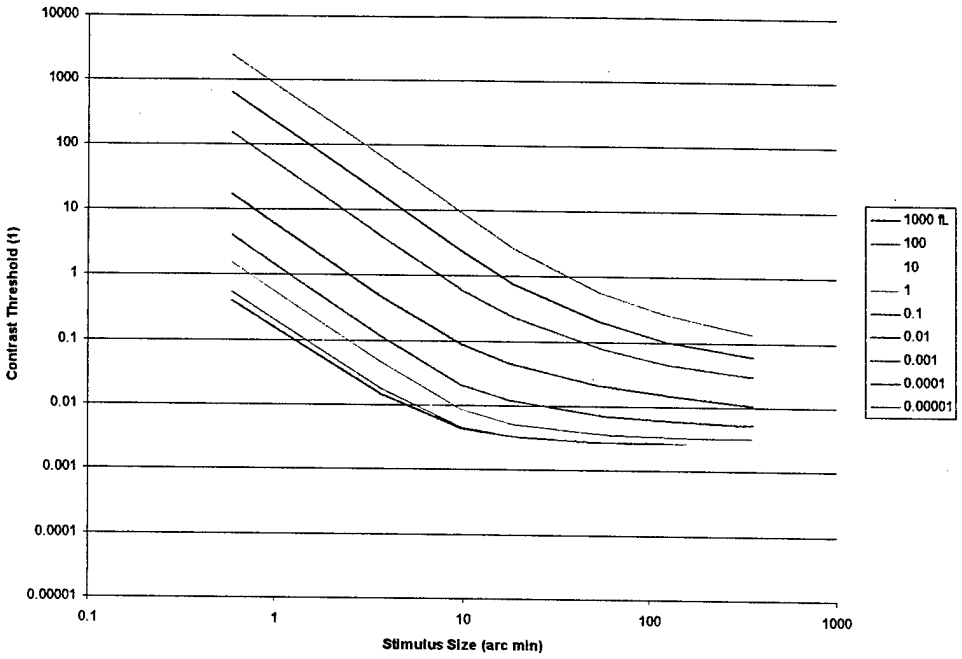


Figure 5a. Contrast threshold vs. stimulus size parameterized by background luminance (re-plotted from data in Reference 10).

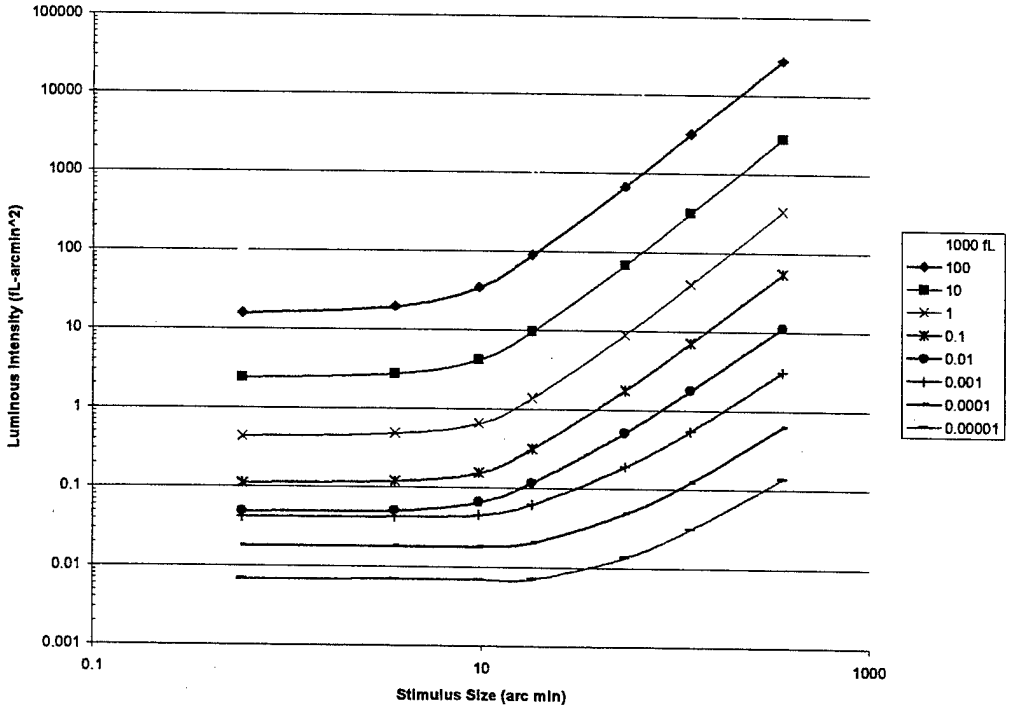


Figure 5b. Luminous intensity threshold vs. stimulus size parameterized by background luminance (re-plotted from data in Reference 10).

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

Figure 6 shows a combined spatio-temporal contrast sensitivity function derived by Kelly, and reproduced in various places (References 2 and 9). This function can be viewed as an effective spatio-temporal MTF for the eye and its peak noise-equivalent contrast, at one illumination level. Note however, that the spatial MTF is based on measurements of linear (one-dimensional) spatial gratings, so that the true two-dimensional spatial MTF is a circularly symmetric function, whose radial profile is described by the spatial function of Figure 6. In theory at least, this combined spatio-temporal MTF function can be thought of as arising from a single spatio-temporal impulse response function describing the eye's complete response to contrast ratio. While many previous models have attempted to build such a comprehensive single channel model, such a model can never be credible, for a series of reasons to be discussed below. However, it is instructive to discuss the steps to be taken in the development of a single channel model, to provide a rationale for the development of the channels making up NAC-VPM.

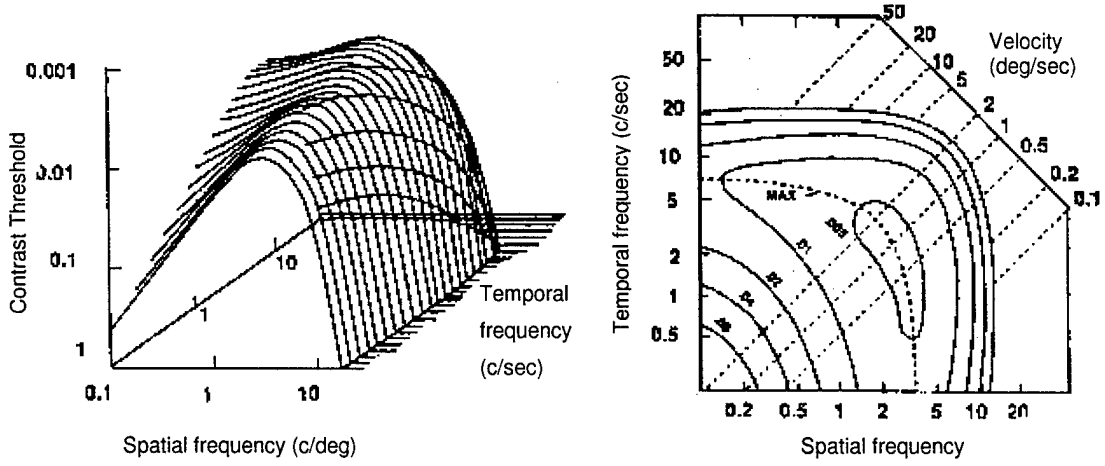


Figure 6. A perspective view of a typical spatiotemporal threshold surface for drifting gratings (left). Each curve represents the spatial frequency response at a fixed temporal frequency. The neighboring curves are separated by a constant increment of about 0.10 log temporal frequency. The hidden part of the surface was not suppressed. A contour map of the same surface (right), the contours are labeled by the contrast thresholds to which they correspond (from Reference 29).

5.3.1 The Attractions of a Linear System Model for Target Detection

Models for human contrast sensitivity based on contrast ratio and functions describing contrast sensitivity have been very successful. The success of these models indicates that, at least for small signals, the eye acts as a linear system. While much has been made of non-linearities in the human visual system, the non-linearities found in the human visual system are primarily those associated with signal detection, such as absolute value functions and

operations such as squaring and raising to a power. The descriptions of pooling are very good examples of signal detection processes (Reference 21). In every sensing system, a non-linear detection process terminates the linear portions of the system. The human visual system appears to be no different than other sensing systems in this respect.

Given the success of linear system theory in modeling the liminal sensitivity of the eye to contrast ratio, one would expect that linear system theory would shed light on the performance of the human visual system in more complex detection processes as well. On the other hand, one might expect that a linear system model might over- or under-predict visual system performance under some conditions. For example, if detection involves a great deal of learning what real targets look like, linear system models might under-predict actual detectabilities. Alternatively, if the eye is not well evolved for certain types of detection, linear system theory might over-predict detectability. We believe, however, that NAC-VPM does not currently optimally predict detectability in a linear system sense.

5.3.2 Possibilities for a Single Channel Linear Model

The simplest computational vision model for target acquisition would be a linear single channel model of dimensionless contrast in terms of total (over the target) contrast ratio signal-to-noise. Such a model would be able to demonstrate all of the features of the eye's luminance contrast sensitivity in both space and time. This model would estimate the eye's ability to detect a complex target in terms of a traditional power signal-to-noise ratio. Many attempts to model detection performance can be viewed as single channel linear models. Metrics such as PSS, four bar pattern models, and models based solely on size and contrast can be viewed as single channel linear system models. Yet there are great disadvantages to such single channel models. Primarily, they fail to account for any ability of the eye to partition the spatio-temporal-color space based on optimum signal-to-noise values. They agglomerate all signal and all noise together into a single channel. Thus they preclude the model from treating any of the advantages associated with selective spatial and temporal frequency filtering and tuning, and from treating any of the advantages that could be gained using the techniques of matched filtering. Thus, one would not expect great success from single channel linear models. However, an outline of the steps to be taken will serve to describe the development of a linear model for any channel whatever, not just a single, all-encompassing channel.

A single-channel signal-to-noise model would be built as follows. The 'signal' would consist of the integral of the square of the perceptible contrast ratio image over the target. This perceptible contrast ratio image can be computed in the following steps:

(1) A luminance adaptation level must be computed for each point in the image, based on a local integral of the scene luminance.

(2) The contrast ratio must now be computed at each point in the image.

(3) If temporal response is to be included, the single image must be replaced by a sequence of images, spanning the time of interest, and extending over twice the duration of the eye's temporal impulse response. Each image in the sequence must be converted into a contrast ratio image.

(4) Now, by convolution, the eye's impulse response function, in both space and time, is applied to each point in the contrast ratio image sequence. The result is a single perceived contrast ratio image. It incorporates both spatial and temporal effects.

(5) The integral of the square of this image must now be taken over the whole of the target-containing region. This becomes the 'signal power' of the channel. The squaring process simulates the 'detection process' necessary in any linear sensing system.

The 'noise power' must now be computed. The simplest model for the 'noise power' contributed by the human eye would simply be the square of the contrast ratio threshold, at the average illumination level associated with the target region, summed over all independent visual spatial resolution elements encompassing the target region. This simple noise model would account for the eye noise generated in a target-sized region.

5.3.3 Background Noise Power in a Single Channel Model

Now the problem of the background comparison must be dealt with. Is the eye actually comparing the target with its associated background, or is the eye viewing the target in isolation, absent from the interfering effects of the background? In the former case, the background is a comparison object with which one is comparing the target, rather than a realization of a random noise process. The background is not a contributor to the noise power in the denominator. In contrast, in the latter case, the background is a noise process, part of the noise against which one tries to detect the target. In the target detection problem, either role can be assigned to the background, but both roles cannot be invoked at once.

Note, however, that while the first choice is almost instinctively chosen, the second choice in the question above is actually the more philosophically appealing. Since there can be more than one realization of a given background against which the eye can be making its comparison, it

is more stochastically reasonable to treat the background as one realization of a confusing process, and so treat it as part of the noise.

In NAC-VPM, the background is assigned a noise-producing role. The signal is that of the target standing alone, and the background is nothing other than a single realization of a confusing noise, within which the target must be sensed. Accordingly, the signal is the square of the dimensionless contrast image of the target itself, without comparison with the surrounding background. The expected value of the square of the signal of the surrounding background (the clutter) becomes a noise component, to be squared and added to the dimensionless contrast threshold noise appearing in the 'noise power' denominator.

Because of the peculiarities associated with multi-resolution spatial filtering, NAC-VPM uses a special image to isolate the target completely from its background. This is the so-called 'bias image'. It is discussed in detail in Reference 1. However, it is not a realization of the background with which the target is compared. It is specially constructed to isolate the target from its background in the multi-resolution images.

If the first alternative is chosen, i.e. the role of the background is taken to be an object with which the target is to be compared, then the signal must be constructed in terms of this comparison. In this case, the actual signal should be the difference between the dimensionless contrast image of the target and an equivalent image of the background to which it is being compared, and the 'signal power' becomes the square of this difference. In this case, the 'noise power' in the denominator contains only the eye noise term. Although this is a valid alternative role for the background, it is not the role used in NAC-VPM.

5.3.4 The Realization of a Single Channel Linear Model

We have now discussed both the signal and noise models for a single channel linear system model for target detection. But there are difficulties in realizing some elements of such a model.

5.3.4.1 The Incorporation of Color

The first, and simplest, difficulty is the incorporation of color into a combined contrast ratio measure. There are several options by which this can be accomplished. It could be done most simply using one of the unified lightness/chromaticity scales used by CIE (Reference 22). The best of these is the L*, U*, V* system defined in Reference 23. Such an extension would

replace the dimensionless luminance contrast ratio based on Equation 2 above with a combined luminance/color contrast ratio based on the L*, U*, V* combined color/lightness scale. This scale was developed based on subjective observer tests relating luminance contrast with color contrast in reflected light, and has been used widely to relate small changes in color to small changes in luminance. It has been heavily used in the printing industry to model the combined perception of brightness and color. It results in a three-element contrast vector ΔL^*, ΔU^*, ΔV^*. This 3-D contrast vector can be combined into a single unified, consistent, contrast value representing both color and luminance differences. It is based on the CIE XYZ color channels.

An alternative approach would make use of the color-opponent chromaticity space of Derrington-Krauskopf-Lennie (DKL) as discussed in References 24 and 25. The DKL space is also a three-element contrast vector, describing contrast in terms of a luminance coordinate, and two color coordinates (the L&M and S coordinates, separated in terms of the three cone spectral sensitivities). While Reference 24 provides substantial discussion about a combined distance measure, no comprehensive, consistent distance measure for this space is described. However, when projected onto the standard CIE chromaticity diagram, the axes of the DKL space are very similar to axes of the L*, U*, V* space. We would thus expect the DKL space to perform similarly to the L*, U*, V* space.

A means to combine luminance and color metrics into a single contrast metric is important to a detection model, since observation of a given target/background image results in a single detection decision (target or no-target), not multiple detections in each of multiple colors. NAC-VPM skirts this issue. It uses independent metrics for both luminance and color channels. The combination of color and luminance effects is controlled both by the relationship between the contrast sensitivity functions for color and for luminance and by the user as he assigns weights to the individual channel metrics. The default weights for all color channels are unity in NAC-VPM, so that the luminance and color channels are always weighted according to their respective contrast sensitivities.

5.3.4.2 The Determination of the Overall System Impulse Response Function

There is a second, much more serious problem with the development of a single channel linear model for detection. It can immediately be realized by noting that the contrast sensitivity (CSF) functions used so commonly to measure the eye's contrast sensitivity are only system response functions. That is, they describe the amplitude of the response to a sine wave input as a function of spatial and temporal frequency. As such, they are inadequate to specify the actual impulse response of the human visual system. Any attempt to build impulse response functions

must immediately address the ambiguity between the impulse response and the system function. Any number of impulse response functions can yield the same system function. One could attempt to resolve this issue by choosing a single even and a single odd function to represent two independent single channels describing the eye's overall sensitivity. However, this immediately breaks what was a single channel into two independent channels with the same system function. The eye's overall system function (represented by the contrast sensitivity function typified by Figure 6) may indeed be the result of the combined action of many independent channels, which are combined in some unknown way to yield an overall detection decision.

The eye's measured system function (the contrast sensitivity function and its variations with color and spatial and temporal frequency) could indeed be the result of a multiplicity of channels, rather than one or a few channels. This realization immediately raises the question of channel selection. Would we not expect the eye to be capitalizing on the variations in signal-to-noise in the spatial-temporal-color domain in which it works? The eye might even dynamically call upon, or construct channels based on the target detection task at hand. We should thus abandon single channel models as being inadequate to model the ability of the eye to recognize objects in images.

5.3.4.3 The Choice of Multiple Channels

If we accept the presence of multiple channels in the human visual system, the problem of modeling detection becomes much more complex. It is no longer possible to envision a single signal-to-noise expression as being capable of modeling target detection. Rather, it becomes necessary to determine a signal-to-noise expression for each of these channels, and then to model the mechanism whereby these multiple channels are combined into a single decision.

How then should these channels be chosen? There is little concrete guidance toward an answer. De Valois and de Valois describe the variety of spatial channels found in primate visual systems in great detail (Reference 9). From their discussion it is clear that a great variety of spatial, temporal and color channels indeed exist in the human visual system. The spatial channels tend to be 'bandpass' channels, i.e. they cover a limited range of spatial frequencies, not extending down to DC, and may or may not extend to the spatial frequency limits of the human visual system. From the discussions in Reference 9, a variety of spatial frequency bandwidths are found, and no particular ordering of the spatial filters is readily apparent in any given visual system. Furthermore, the observed spatial size of the receptive fields may be unrelated to the observed size of their central lobes. The lack of any fixed relationship between receptive field size and central lobe size, and the lack of observed order in the receptive fields, are both strong indications of the presence of a great variety of receptive fields. They may have many different

(and probably unrelated) central frequencies and spectral resolutions (the ratio of bandwidth to central frequency).

A similar situation exists for temporal channels. While the overall temporal contrast sensitivity function extends from DC to perhaps 30Hz, observed temporal channels show much narrower bandwidths, indicating that the temporal response is the summed response of several temporal channels.

5.3.4.4 NAC-VPM/TVM Channel Choices

The approach used to define multiple channels in NAC-VPM has been inherited completely from the older code, TVM.

For the details of these channels, the reader is referred to the NAC-VPM report and the TVM Analyst's Manual (References 1 and 2). The choices made for these two codes were made based on the best estimates of vision channelization available at the time TVM was written.

The temporal channels in TVM, and hence in its successor, NAC-VPM, are based on the work of Kelly (Reference 21). Kelly describes a temporal contrast sensitivity function in detail, reproduced here as Figure 7. For the TVM effort (and inherited by NAC-VPM) we divided this contrast sensitivity function arbitrarily into three channels, one a temporal lowpass channel and the other two temporal bandpass channels. The system functions of these three channels are such that their sum duplicates that of Kelly. (See Figure 7.) When implemented in the TVM/NAC-VPM temporal preprocessor, the system functions of these three filters were taken to be real and symmetric about the frequency origin. This means that their impulse response functions have been chosen to be even functions. Thus there are three temporal channels in TVM/NAC-VPM, which sum to match the contrast sensitivity function of Kelly. Their impulse responses are all even. Note, however, that we could as easily have chosen odd functions, or any mixed set of even-odd functions that satisfied Kelly's contrast sensitivity function.

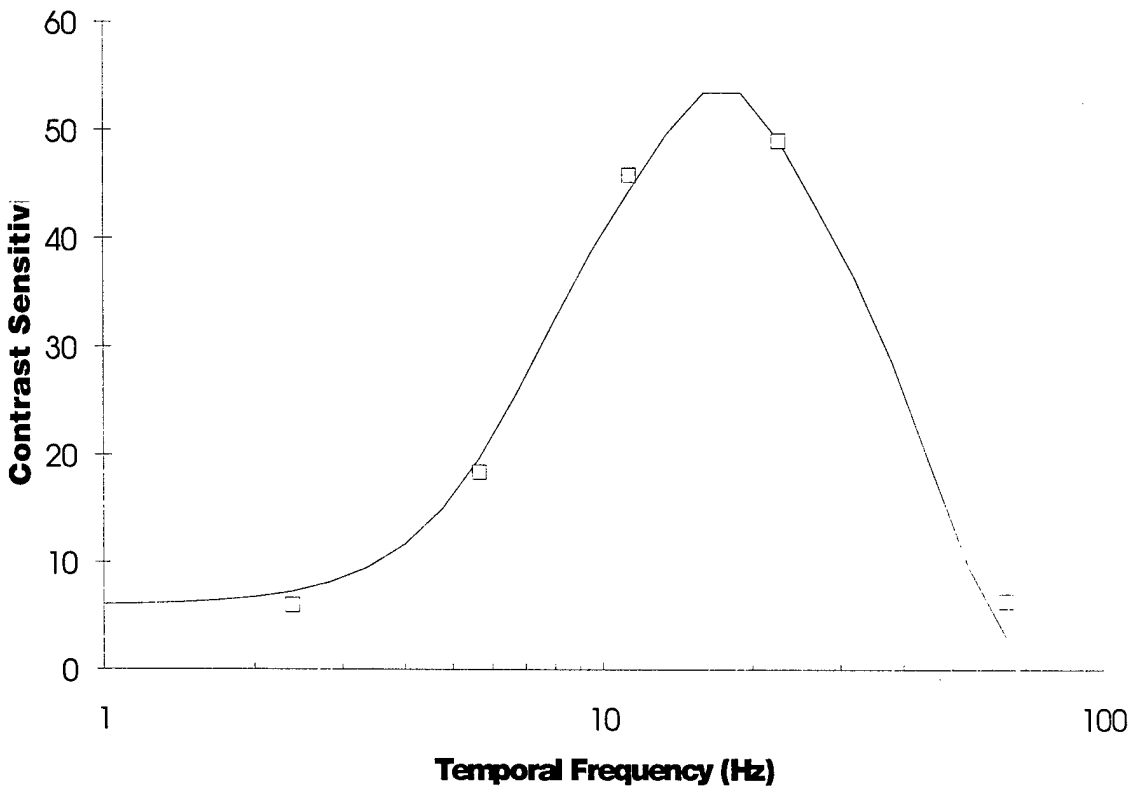


Figure 7. NAC-VPM's temporal contrast sensitivity model (curves) and relevant data points at five luminance levels (Reference 26).

NAC-VPM/TVM develops both lowpass and bandpass temporal channels for the luminance image sequence only, and develop only a lowpass temporal channel for the color opponent image sequences – thus the codes assume that the human vision system has no color vision in the temporal bandpass channels.

Each of the three luminance temporal channels and the two color opponent channels are divided into a number of spatial channels. This division into spatial channels has been based on the hierarchical pyramidal representation schemes of Burt and Adelson (Reference 27). In this approach, each image is subdivided into a pyramid of images, each of lower spatial resolution than the one above it in the pyramid. The frequency range covered by each image in the pyramid spans the outer one octave wide ring of spatial frequency space, for that image, as shown in Figure 8. Thus, through the use of the Burt and Adelson pyramid, the full spatial frequency space of the original image is covered with channels about one octave in bandwidth.

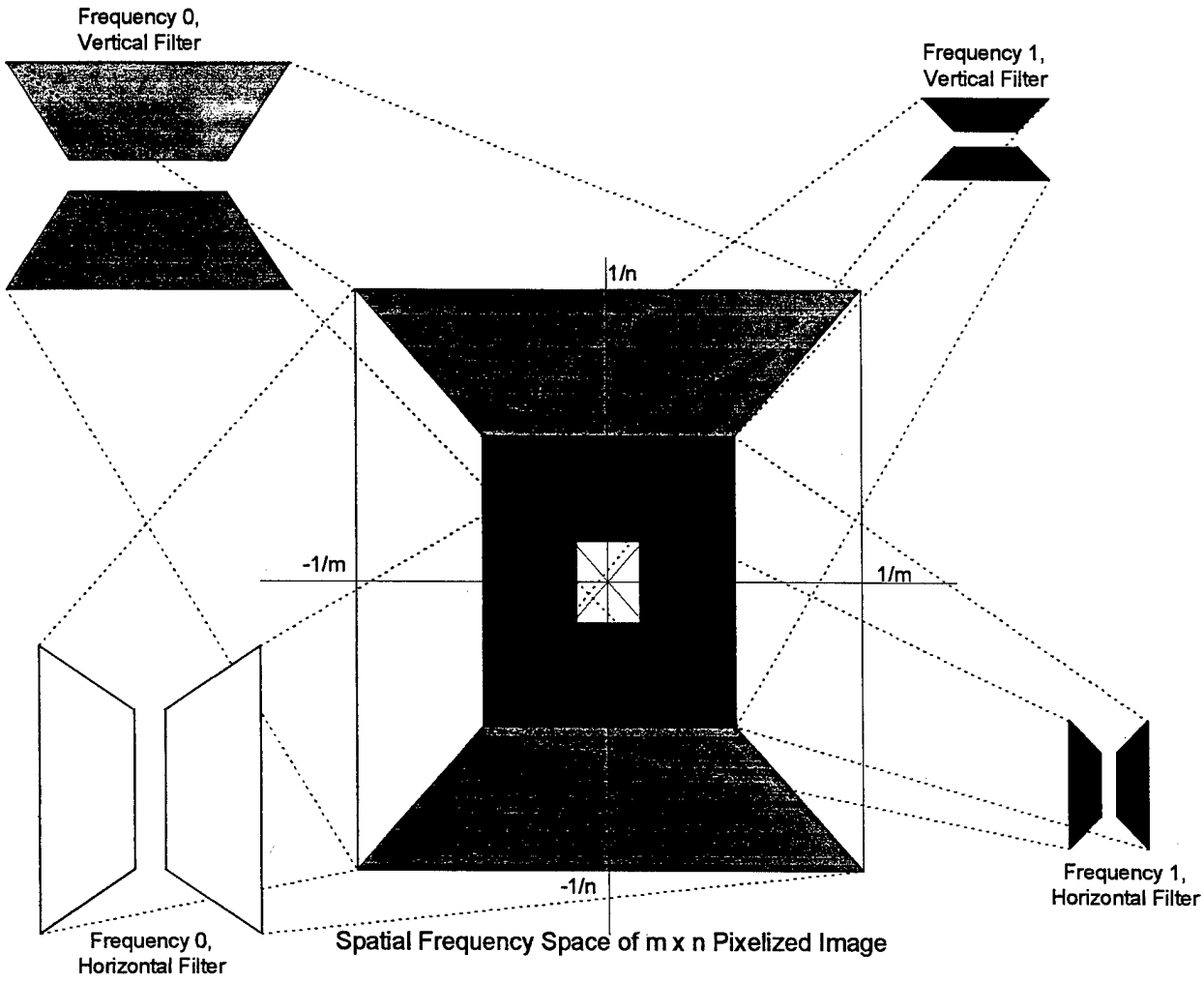


Figure 8. Partitioning of spatial frequency space by NAC-VPM's directional spatial frequency pyramid. Partitioning is shown for two outermost rings of frequency space. Each successive ring is partitioned similarly.

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

Finally, the color separation is based on luminance and two color opponent channels described by Boynton (Reference 28).

Thus, the channels chosen for NAC-VPM can be summarized as follows: The temporal channels are based on the temporal contrast sensitivity of Kelly. The spatial channels are chosen based on the hierarchical decomposition of Burt and Adelson. Color separation is according to luminance and Boynton's two color opponent channels. Color, spatial and temporal channelizations are all assumed to be independent. Note that none of the channel choices are based explicitly on any observed channels in the human visual system. Indeed, given the variety of channels observed and discussed in Reference 9, it would be difficult to pick any particular channels based on observed receptive fields.

These choices have lead to several unfortunate outcomes. First, the spatial channelization is fairly coarse, each channel being an octave wide, even though as many as 7 octaves of spatial frequency space is spanned. Furthermore, the channelization is based not on any characteristics of the target or human visual system, but on the pixelization of the image itself. This is particularly unfortunate, since there is no guarantee that the target's spatially important information is isolated into one or more particular channels for consideration by the visual system. The channels are very wide and very sparsely chosen. The only virtue associated with this spatial choice is that all of the spatial frequency space is covered.

A second outcome is that, because of the scheme used to implement the hierarchical decomposition, all of the channels' impulse response functions are 'even' functions in both spatial dimensions. This is unfortunate since such a set of filters cannot preserve phase information. Preservation of phase is known to require channels having both 'even' and 'odd' impulse responses. Yet the preservation of phase is known to be critical for object recognition. Thus, the pyramidal representation as implemented cannot preserve phase information associated with the bandpass filters. The temporal filters are also limited to even functions, although this may not be a problem since there is little evidence that temporal phase is important for moving target detection.

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

In summary, the channel selections made during the design of TVM, as inherited by NAC-VPM, are somewhat arbitrary, based on a technique designed for image compression (the Burt and Adelson hierarchical pyramid), and on the overall temporal contrast transfer function. In particular, they make no pretense of being chosen based on either an actual visual channelization, nor on actual target spatial frequency content. Based on these observations, the imperfect correlation of results from these models with real observer performance may not be unexpected.

5.3.4.5 How Should Channels be Chosen?

Now that NAC-VPM is complete, the channel choices can be critiqued in hindsight. When channels were chosen, the features leading to the criticisms above were considered to be virtues. Thus the use of the temporal contrast sensitivity function as the basis for the selection of three temporal channels was considered as valid. At the time, little consideration was given to the difference between a system function (a representation of amplitude response) and the actual impulse response for a channel.

Furthermore, the spanning of all spatial frequency space, and the ability to reconstruct the image from the pyramid was valued above the preservation of phase information, and little consideration was given to the details of the channel choices. These choices have almost certainly compromised the ability of NAC-VPM/TVM to successfully model target detection problems. We can conceive of better ways to choose channels.

We could postulate a human visual system (HVS) which works approximately as follows: the HVS could be blessed with a great variety of spatial, temporal and color channels which serve as programmable tools for its higher level cognitive processes. Each task given to the human visual system, whether it be a complex target detection task, or a simple liminal contrast perception task, may be handled at the cognitive level. The cognitive level of the visual system then calls upon its repertoire of channels to analyze the perceived image, resulting in a decision. The ultimate capability of such a system would still be describable in terms of a complex contrast sensitivity function, describing the ultimate capabilities of its low level sensors. Nevertheless, its actual complex detection capabilities would be measurable only in the characteristics of its cognitive processes.

The investigations described in Section 3 give some support to this method of operation. The physiological studies described there show the existence of a great variety of receptive fields, organized in several ways. But no structures supporting particular high level tasks, such as brightness matching, can be found. Further evidence can be gained from descriptions of a variety of spatio-temporal-color channels in Reference 9, combined with no real evidence for the

existence of processes that analyze brightness levels or which are specific to particular textures. Perhaps all decisions are handled at a very high level based on the considered results from a large repertoire of perhaps even programmable spatio-temporal-color channels. This is in contrast to NAC-VPM and its predecessor, TVM, in which the channels are fixed, small in number, and chosen based on scene pixelization rather than on the target.

It is possible to choose channels based on the postulation described above. First of all, NAC-VPM could have a much richer choice of channels. A large repertoire of channels could be available, including various selections of relative bandwidth, and central frequencies matched to particular feature sizes. Both even and odd impulse responses should be available for each channel. Selections from this repertoire can then be made at the time of metric calculation, based on characteristics of the target such as its size, its range, its texture or lack thereof, and its probable relationship to its background. Linear system signal-to-noise metrics would be computed for every channel selected from the repertoire. The contrast sensitivity functions would still be used as a source of eye noise values for all channels. Then, rather than simply combining the signal-to-noise metric from all selected channels into one decision-level metric, the channel signal-to-noise ratios would be sorted by value, and the overall detection metric then computed as a combination of metrics from those few channels with the largest signal-to-noise ratios.

If we were to postulate, in addition, that the human visual system can actually implement the spatio-temporal filters required to detect any realistic target, we can then build a detection metric which is even more efficient. We can perform an eigenvalue-eigenvector analysis of the target region and find those spatio-temporal filters which contain the most target energy. We can then postulate the existence of those channels and compute signal-to-noise metrics for them. These metrics can then form the basis of a detection model.

These schemes for choosing channels for consideration have much more appeal than the fixed set of channels presently used in NAC-VPM. Thus, to the extent that effort in visual detection modeling is continued, primary attention should be given to selection of visual channels.

6. Conclusions and Recommendations

The primary results of this program consist of (1) the completion and delivery of the NAC-VPM software, (2) the validation of the software against a 'conspicuity' experiment involving the identification of oncoming traffic by a visual observer. This validation effort was moderately successful, in that the NAC-VPM model yielded correlation of about 0.8 with actual observer tests.

In addition, several additional efforts were undertaken directed toward more complete understanding of the visual detection problem. The results of these efforts were somewhat inconclusive. While interesting transformations of image data were developed, much more investigation remains to be done before any substantive conclusions about them can be reached.

Finally, a critique was made of the philosophy behind linear detection models, of which NAC-VPM is a prime example. This critique demonstrated that the channel selections made in the development of NAC-VPM are not optimum for the detection of the targets being investigated. Future efforts should be directed toward the development of techniques, which select visual channels appropriate to the detection task at hand.

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

7. References

1. Witus, G., TARDEC National Automotive Center Visual Perception Model (NAC-VPM) – Final Report: Analyst's Manual and User's Manual, OMI-577, OptiMetrics, Inc., 1996.

2. Lindquist, G. et. al., TARDEC Visual Model Version 2.1.1 Analyst's Manual, OMI-552, OptiMetrics, Inc., 1995.

3. Smith, F. and A. Dunstan, Visual Performance Model Analysis of Human Performance in IR Recognition, OMI-608, OptiMetrics, Inc., 1997.

4. Tootel, et. al., Journal of Neuroscience, 8(5), May 1988.

5. Croner,L. and E. Kaplan, 'Receptive Fields of P and M Ganglion Cells Across the Primate Retina,' Vision Research, 35:1, 1995.

6. Eckhorn, R., F. Krause and J. Nelson, Biol. Cybern. 69, 37-55, 1993.

7. Tovee, M.J. An Introduction to the Visual System, Cambridge University Press, 1996.

8. Roe, A. W. and Ts'o, D., Visual Topography in Primate V2: Multiple Representation Across Functional Stripes, Journal of Neuroscience, 15, 3689-3715, 1995.

9. R. L. De Valois and K. K. De Valois, Spatial Vision, Oxford University Press, 1990.

10. Blackwell, H. R., Contrast Thresholds of the Human Eye, Journal of the Optical Society of America, 36:11, 624-643, 1946.

11. Horn, B. K. P., Robot Vision, MIT Press, 1986.

12. Eisenhart, L. P., An Introduction to Differential Geometry, Princeton University Press, 1947.

13. Landau, L. D. and E. M. Lipshitz, The Classical Theory of Fields, Pergamon Press, 1962.

14. Gabor, D. 'Theory of Communication,' Journal of Institute of Electronic Engineers, Vol 93,Part III, ab26, 951-966, 1946.

15. Berry, M. V., Proc. R. Society, A392, 45, 1984.

16. Bohm, A., Quantum Mechanics: Foundations and Applications, Springer-Verlag, Third Edition, 1994.

17. Page, Don N., Geometrical Description of Berry's Phase, Phys. Rev. A, 36:7, 3479, 1987.

18. Dudzak, M. C. editor, The Infrared and Electro-Optical Systems Handbook, Vol. 4, 'Electro-Optical Systems Design, Analysis and Testing, Chapter 2, 'Electro-Optical Imaging Systems Performance Prediction,' ERIM, Ann Arbor, MI, and SPIE, Bellingham, WA, 1993.

19. Kelly, D. H., 'Receptive Field-Like Functions Inferred from Large-Area Physchophysical Measurements,' Vision Research, 25:12, 1985. Printed in Great Britian.

20. Kornfeld, G. H. and W. R. Lawson, 'Visual-Perception Model,' Journal of the Optical Society of America, 61:6, 1971.

21. Graham, N. 'Complex Channels. Easy Local Non-linearities and Normalization in Texture Segregation' in Computational Models of Visual Processing, M. S. Lordy and J. A. Morston, ed., MIT Press, Cambridge, MA, 1991.

22. Wyszecki, G., 'Colorimetry,' in Handbook of Optics, W. A. Driscoll and W. Vaughn, ed., McGraw Hill, 1978.

23. Wyszecki, G. and W. S. Stiles, Color Science, Concepts and Methods, Quantitative Data and Formulas, Wiley and Sons, 1967.

24. Kaiser, P. K. and R. M. Boynton, Human Color Vision, 2nd Ed., Optical Society of America, 1996.

25. Derrington, A. M., Krauskopf, J., Lennie, P., 'Chromatic Mechanisms in Lateral Geniculate Nucleus of Macaque,' Journal of Physiology, 357, 244-265, 1984.

26. Kelly, D. H., 'Visual Responses to Time-Dependent Stimuli, I. Amplitude Sensitivity Measurements,' Journal of Optical Socity of America, 51, 522-529 1961.

27. Burt, P. J. and E. H. Adelson, 'The Laplacian Pyramid as a Compact Image Code,' IEEE Transaction on Communications, CON-Vol 31, No. 4, 1983.

28. Boynton, R. M. Color Vision, Optical Society of America, 1992.

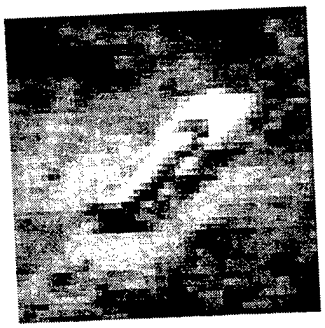
A. Appendix A

During the past year, several problems with NAC-VPM were noted and repaired. These repairs have been included in the last distribution of NAC-VPM and are described below.

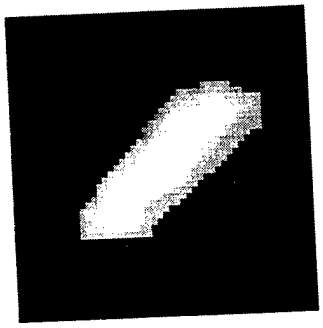
Problem with Bias Image Computation

Problem Description

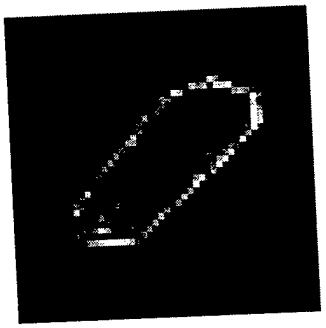
While running some test cases it became apparent that the code was not working as expected. Nearly of all the VPM metric contributions were from the highest frequency channel. Upon inspection of the images (TargetRFDetectability) from which the metric is calculated, the target appeared with a strong outline and little internal detail. However, this was in conflict with the character of the input image, which had relatively weak boundary contrast but considerable internal detail. The TargetRFDetectability images are derived from the difference between the (processed) input image and a BiasImage. The BiasImage is created from the input image, but with the target removed and the surrounding background 'blended-in' to fill the area previously occupied by the target. Upon inspection, the BiasImage was also found to contain the same problematic characteristics. The source of errors in the TargetRFDetectability images and thus the metric results were traced to an error in the computed BiasImage. Examples of the input and of the erroneous intermediate images are shown below.



Input Image



Blended Background Image



RFTargetDetectability Image

Figure A.1. Example images.

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

Problem Solution

The key calculations that create the BiasImage are defined in the file, omiBlendOut2.map. This map implements much of the processing described in pages 17-27 in Reference 1. As an initial step in the processing, an offset (actually called a bias constant) is added to the image. Later in the processing that offset is removed. Halfway through the processing in omiBlendOut2, an intermediate image, called Image1 below, is created which improperly includes the offset. There is also blending function, NonStationaryLowpassInverse, (NLSI below). The last lines of omiBlendOut2 produce the following calculations:

NSLI{Threshold(Image1) * Image1 } - Offset

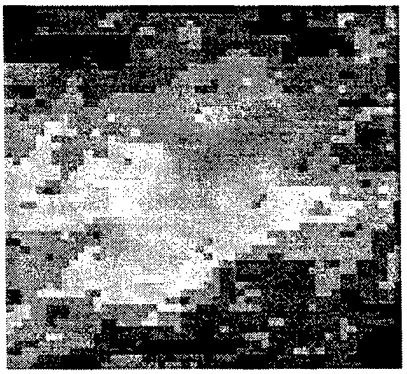
As shown, the NSLI process is applied before the offset is removed. That results in extrapolation of erroneously large background components into the target area of the image. When the offset is later subtracted it results in an outline of the target area.

To correct this problem, the processing was modified to execute as follows:

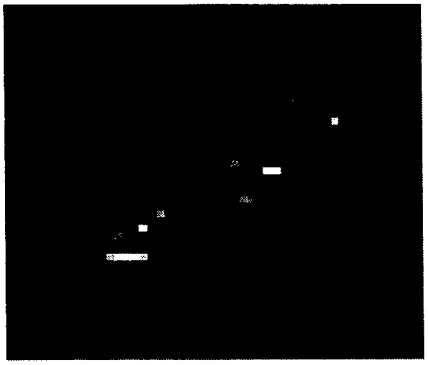
NSLI{Threshold(Image1) * Image1 – Offset}

In this case, the offset is first subtracted and then the non-linear blending operation is performed.

This modified operation is incorporated in a modified version of the omiBlendOut3.map file. Examples of the RFTargetDetectability and BiasImage images created with the modified processing are shown below. Both have the expected character, with the BiasImage showing a smooth blending from the background into the target region and the RFTargetDetectability image showing a mixture of edge and internal features.



Bias Image



RFTargetDetectability Image

Figure A.2. Example images after processing correction.

Errors in Color Transformation

Problem Description

When NAC-VPM was used to model detectability in black and white images, spurious color opponent channels appeared. Furthermore, color images were improperly transformed.

Problem Solution

The problem was traced to an inappropriate RGB to XYZ transformation matrix, and its corresponding gamma correction parameters. This transformation data is part of the default data provided with NAC-VPM. To prevent the problem from occurring with standard RGB images, the default transformation was replaced with a more appropriate transformation. Before NAC-VPM can give accurate outputs, this transformation must be replaced with the actual transformation between the RGB image in use and the luminance in XYZ space at the observer.

There are constraints to this transformation in that it must not introduce color when no color is indeed present. Furthermore, it must account for the white point of the display and for the colors of the phosphors. To reproduce the white of a CIE standard C illuminant, the sums of the three columns of the matrix should be in the ratio (0.981, 1, 1.1836). Furthermore, the chromaticities of each row should match the chromaticity points of the three phosphors. The matrix below meets all these constraints. It simulates a linear RGB display with a maximum brightness of 255 cd/m2, and whose RGB channels match the standard R, G, B phosphor chromaticities of (0.62, 0.34), (0.28, 0.59) and (0.15, 0.07), respectively. The sum of the center column is unity, so that, to give a simple calibration, it can be multiplied by a constant equal to the

Contract No. DAAE07-96-C-X053
Contractor OptiMetrics, Inc.
Address 3115 Professional Drive
Ann Arbor, Michigan 48104
Expiration of SBIR Data Rights Period March 13, 2003

luminance of a gray image of value 1 (i.e. R = G = B = 1).

	X	Y	Z
R	0.480	0.263	0.031
G	0.307	0.646	0.142
B	0.194	0.091	1.010

This transformation is located in the file COAInputData, and appears as the RGBtoXYZ matrix in this file, along with default bias and exponent values.

The Addition of Spatial Weights

Description of Problem

During evaluation of certain images, it became apparent that the highest frequency channels often consist of nothing but white noise. Yet, because of their large variance, they dominated the signal content.

Problem solution

To allow the user control over situations of this type, provision was added to assign a user-defined weight to each spatial channel in each of the five color planes. This was implemented in the form of an additional set of data files named TLBW/TLRG/TLYB/TMBW/THBWSpatialWeights. These files are in the same format as all other data files. The metric processors now properly recognize these files and use the weights contained in them. The default files having these names all have all the weights set to unity.

B. Appendix B -- Bibliography

The following table of references describes all sources of information used in the supplementary studies described in sections 2 through 5.

AUTHOR	TITLE	REFERENCE	
Ahumada, A.J., JR., A. B. Watson, A. M. Rohaly	'Models of human image discrimination predict object detection in natural backgrounds'	*SPIE, Vol. 2411/355*	1-6, 355-362
Akerman, A. III	'Predicting aircraft detectability'	*Human Factors, 21(3), 1979*	277-291
Albrecht, D. G. and Hamilton, D. B.	'Striate cortex of monkey and cat: Contrast response function'	*Journal of Neurophysiology, Vol. 48, No. 1, 1982*	217-237
Anstis, S. M.	'A chart demonstrating variations in acuity with retinal position'	*Vision Research, Vol. 14, 1974*	1-82
Antoine, J. P., P. Carette, R. Murenzi and B. Piette	'Image analysis with two-dimensional continuous wavelet transform'	*Signal Processing, 31 (1993)*	
Ballard, D. H.	'Generalizing the Hough Transform to Detect Arbitrary Shapes'	*Pattern Recognition, Vol. 13, No. 2., 1981*	
Becker, S.	'Mutual information maximization: models of cortical self-organization'	*Network: Computation in Neural Systems, 7(1996) UK.*	589-593
Bergen, J.R. and Landy, M.S.	'Computational Modeling of Visual Texture and Segregation'		241-272
Bergen, J.R., Wilson, H.R., and Cowan, J.D.	'Further Evidence for Four Mechanisms Mediating Vision at Threshold: Sensitivities to complex gratings and aperiodic stimuli'	*Journal of the Optical Society of America, Vol. 69, No. 11, 1979*	
Blackwell, H.R.	'Contrast thresholds of the human eye'	*Journal of the Optical Society of America, Vol. 36, No. 11, 1946*	i-89
Blackwell, H.R.	'Neural theories of simple visual discriminations'	*Journal of the Optical Society of America, Vol. 53, No. 1, 1963*	111-122,

Bialek, W. and Zee, A.	'Understanding the efficiency of human perception'	*Physical Review Letters, Vol. 61, No. 13, 1968.*	
Bloomfield, J.R.	'Visual search in complex fields: Size Differences between target disc and surrounding discs'	*Human Factors, Vol. 14, No. 2, 1972*	
Bouchiat, C. and Gibbons, G. W.	'Non-integrable quantum phase in the evolution of a spin-1 system: a physical consequence of the non-trivial topology of the quantum state-space'	*J. Phys., 49, 1988, France*	7-31
Boulton, J.C.	'Two mechanisms for the detection of slow motion'	*Journal of the Optical Society of America, Vol. 4, No. 8, 1987*	
Braccini, C., G. Gambardella, and G. Sandini	'A Signal Theory Approach To The Space And Frequency Variant Filtering Performed By The Human Visual System'	*Signal Processing, Vol. 3, 1981*	
Bronskill, J.F., J.S.A. Hepburn, Au K. Wing	'A Knowledge-Based Approach to the Detection, Tracking and Classification of Target Formations in Infrared Image Sequences,'	*IEEE, 1989*	624-632
Burt. P.J., and Adelson, E.H.	'The Laplacian Pyramid as a Compact Image Code,'	*IEEE Transaction on Communications, CON-Vol. 31, No. 4, 1983*	
Cannon, M. W.	'A transducer model for contrast perception'	*Vision Research, Vol. 31, No. 6, 1991*	1512-1515
Cannon, M .W., Jr., and Steven C. Fullenkamp	'Perceived contrast and stimulus size: Experiment and simulation'	*Vision Research, Vol. 28, No. 6, 1988*	139-148
Cannon, .W., Jr.	'Perceived contrast in the fovea and periphery'	*Journal of the Optical Society of America, Vol. 2, 1985*	
Cannon, M. W. and Fullenkamp, S. C.	'Spatial interactions in apparent contrast: Inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations'	*Vision Research, Vol. 31, No. 11, 1991*	187-199

Chellappa, R., Q. Zheng, P. Burlina, C. Shekhar, K.B. Eom	'On the Positioning of Multisensor Imagery for Exploitation and Target Recognition'	*Proceedings of the IEEE, Vol. 85, No. 1, Jan. 1997*	1634-1642
Cohn, T. E., and Lasley, D. J.	'Detectability of a luminance increment: effect of spatial uncertainty'	*Journal of the Optical Society of America, Vol. 64, No. 12, 1974*	91-100
Cohn, T. E. and Wardlaw, J. G.	'Effect of large spatial uncertainty on foveal luminance increment detectability'	*Optical Society of America, Vol. 2, No. 6, 1985*	231-240
Croner, Lisa J., E. Kaplan	'Receptive Fields of P and M Ganglion Cells Across the Primate Retina'	*Vision Research, Vol. 35, No. 1, 1995*	153-157
D'Zmura, M.	'Color in visual search'	*Vision Research, Vol. 31, No. 6, 1991*	532-540
Dannemiller, James L.	'Spectral reflectance of natural objects: how many basis functions are necessary?'	*J. Opt. Soc. Am. A, Vol. 9, No. 4, April 1992*	
Daughman, J. G.	'Two-dimensional spectral analysis of cortical receptive field profiles'	*Vision Research, Vol. 20, 1979*	983-998
Davis, E. T. and Graham, N.	'Spatial frequency uncertainty effects in the detection of sinusoidal gratings'	*Vision Research, Vol. 21, 17 March 1980*	695-709
Davis, E. T., Kramer, P. and Graham, N.	'Uncertainty about spatial frequency, spatial position, or contrast of visual patterns'	*Perception & Psychophysics, Vol. 33, No. 1, 1983*	1760-1768
DeAngelis, G. C., I. Ohzawa, R. D. Freeman	'Receptive-field dynamics in the central visual pathways'	*Trends Neuroscience, (1995) 18*	1985-1998
De Valois, R. L., Webster, M. A. and De Valois, K. K.	'Temporal properties of brightness and color induction'	*Vision Research, Vol. 26, No. 6, 1986*	
Duncan, J. and Humphreys, G. W.	'Visual Search and stimulus similarity'	*Psychological Review, Vol. 96, No. 3, 1989*	
Eastman, A. A.	'Color contrast vs. luminance contrast'	*Illuminating Engineering, Vol. 63, 1968.*	1715-1719

Edwards, D. P., K. P. Purpura, E. Kaplan	'Contrast Sensitivity and Spatial Frequency Response of Primate Cortical Neurons in and Around the Cytochrome Oxidase Blobs'	*Vision Research, Vol 35, No. 11, 1995*	820-825
Forsyth, D.; Mundy, J.; Zisserman A.; Coelho, C.; Heller, A.; Rothwell, C.	'Invariant Descriptors for 3D Object Recognition and Pose'	*IEE Transactions on Pattern Analysis and Machine Intelligence, Vol 13, No. 10*	7-24
Freeman, W. T. and Adelson, E. H.	'The design and use of steerable filters'	*IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9, 1991*	
Gabor, D.	'Theory of communication'	*Journal of Institute of Electronic Engineers, Vol. 93, Part III, #26, Nov. 1946*	951-966
Gaudart, L., J. Crebassa, J. P. Petrakian	'Wavelet transform in human visual channels'	*Applied Optics, Vol. 32, No. 22, 1 August 1993*	507-515
Graham, N., Kramer, P., and Yager, D.	'Signal-detection models for multidimensional stimuli: probability distributions and combination rules'	*Journal of Mathematical Psychology, Vol. 31, 1987*	847-856
Graham, Norma, Beck, J., Sutter, A.	'Nonlinear Processes in Spatial-frequency Channel Models of Perceived Texture Segregation: Effects of Sign and Amount of Contrast'	*Vision Research, Vol. 31, No. 4, 1992*	705-712
Graham, N., A. Sutter, C. Venkatesan, M. Humaran	'Non-linear processes in perceived region segregation: orientation selectivity of complex channels'	*Ophthal. Physiol. Opt., Vol. 12, April 1992*	20-28
Graham, N., A. Sutter, C. Venkatesan	'Spatial-frequency-and Orientation-Selectivity of Simple and Complex Channels in Region Segregation'	*Vision Research, Vol. 33, No. 14, 1993*	451-458
Greening, C. P. and Wyman, M. J.	'Experimental evaluation of a visual detection model'	*Human Factors, Vol. 12, No. 5, 1970*	209-214
Greening, C. P.	'Mathematical modeling of air-to-ground target acquisition'	*Human Factors, Vol. 18, No. 2, 1976*	

Greig, G. L.	'On the shape of energy-detection ROC curves'	*Perception and Psychophysics, Vol. 4 8, No. 1, 1990*	
	Ground Target Modeling and Validation Conference	*August 20-22, 1996- Abstracts*	
Haralick, R. M.	'Statistical and structural approaches to texture'	*Proceedings of the IEEE, Vol. 67, No. 5, 1979*	887-897
Hecker, R.	'Chameleon- Camouflage assessment by evaluation of local energy, spatial frequency and orientation'	*Industrieanlagen-Betriebsgesellschaft m. b. H.*	
Jacobson, L., and Wechsler, H.	'Derivation of optical flow using a spatiotemporal-frequency approach'	*Computer Vision and Image Processing, Vol. 38, 1987*	
Jacobson, L.	'Human Performance Modeling for EOTDAs'	*NM WIDA Meeting, 22-23 March 1994, Las Vegas, NM*	37-110
Jacobson, L., and Wechsler, H.	'Invariant Architectures For Low-Level Vision'	*Computer Vision and Image Processing, 1992*	
Johnson, J.	'Analysis of Image Forming Systems'	*U. S. Army Engineer Research and Development Laboratories, VA*	
Johnson, K. R.	'Ground Target Modeling and Validation Conference'	*Volume II Proceedings, Sixth Annual, August 1995*	iii-35
Judd, D. B. and Eastman, A. A.	'Prediction of Target Visibility From the Colors of Target and Surround'	*Illuminating Engineering, Vol. 66, No. 4, 1971*	433-458
Kelly, D. H.	'Adaptation Effects on Spatio-Temporal Sine-Wave Thresholds'	*Vision Research, Vol. 12, 1972*	613-619
Kelly, D. H.	'Receptive-Field-Like Functions Inferred from Large-Area Psychophysical Measurements'	*Vision Research, Vol. 25, No. 12, 1985, Printed in Great Britain*	1501-1523
Kersten, D.	'Spatial Summation in Visual Noise'	*Vision Research, Vol. 24, No. 12, 1984*	1-106

Koenderink, J. J., Bouman, M. A., Bueno de Mesquita, A. E., and Slappendel, S.	'Perimetry of Contrast Detection Thresholds of Moving Spatial Sine Wave Patterns. IV. The Influence of the Mean Retinal Illuminance'	*Journal of the Optical Society of America, Vol. 68, No. 6, 1978*	
Kornfield, G. H. and Lawson, W. R.	'Visual-Perception Models'	*Journal of the Optical society of America, Vol. 61, No. 6, 1971*	454-460
Kramer, P., Graham, N., and Yager, D.	'Simultaneous Measurement of Spatial-Frequency Summation and Uncertainty Effects'	*Optical Society of America, Vol. 2, No. 9, 1985*	891-906
Kukkonen, H., J. Rovamo, R. Näsänen	'Masking Potency and Whiteness of Noise at Various Noise Check Sizes'	*Investigative Ophthalmology & Visual Science, Vol. 36, No. 2., February 1995*	429-457
Landy, M. S.	REFERENCES	*Computational Models of Visual Processing, 1991*	
Landy, M. S. and Bergen, J. R	'Texture Segregation and Orientation Gradient'	*Vision Research, Vol. 31, No. 4, 1991*	
Lawton, T. B.	'Dynamic Object-Based 3-D Scene Analysis Using Multiple Cues'	*In SPIE Proceedings Computational Vision Based on Neurobiology, Vol. 2054*	1-25
Lawton, T. B.	'Image Enhancement Filters Significantly Improve Reading Performance For Low Vision Observers'	*Opthal. Physiol. Opt., Vol. 12, April 1992*	4119-4127
Lawton, T. B.	'Neural Algorithms for Automated Pattern Recognition in Natural Scenes'	*IEEE Transactions on Parallel Processing, in Fullerton, CA, 4 April 1990*	
Lawton, T. B., C. W. Tyler	'On the Role of X and Simple Cells in Human Contrast Processing'	*Vision Res., Vol. 00, No. 0, 1993*	ii-143
Lawton, T. B.	'Outputs of Paired Gabor Filters Summed Across the Background Frame of Reference Predict the Direction of Movement'	*IEEE Transactions on Biomedical Engineering, Vol. 36, No. 1, January 1989*	
Lawton, T. B.	'The Effect of Phase Structure on Spatial Phase Discrimination'	*Vision Res., Vol. 24, No. 2, 1984*	

Lee, J. C., E. Milios	'Matching Range Images of Human Faces'	*IEEE, 1990*	
Legge, G. E., and Foley, J. M.	'Control Masking in Human Vision'	*Journal of the Optical Society of America, Vol. 70, No. 12, 1980*	136-144
Leventhal, A. G., K. G. Thompson, D. Liu, Y. Zhou, and S. J. Ault	'Concomitant Sensitivity to Orientation, Direction, and Color of Cells in layers 2, 3, and 4 of Monkey Striate Cortex'	*The Journal of Neuroscience, March 1995, 15(3)*	
Li, Hui; B.S. Manjunath and Sanjit K. Mitra	'A Contour-Based Approach to Multisensor Image Registration'	*IEEE Transactions on Image Processing, Vol. 4, No. 3, March 1995*	1-10
Lubin, Jeffrey	'The Use of Psychophysical Data and Models in the Analysis of Display System Performance'	*Visual Factors in Electronic Image Communication, A. B. Watson, ed., MIT Press, 1993*	
Miyahara, E., V. C. Smith, J. Pokorny	'How surrounds affect chromaticity discrimination'	*Journal Opt. Soc. Am. A., Vol. 10, No. 4, April 1993*	366-409
Mundy, J. L., A. J. Heller	'The Evolution and Testing of a Model-Based Object Recognition System'	*IEEE, 1990*	719-743
Näsänen, R. E., H. T. Kukkonen, J. M. Rovamo	'A Window Model for Spatial Integration in Human Pattern Discrimination'	*Investigative Ophthalmology & Visual Science, Vol. 36, No. 9, August 1995*	142-146
Nielsen, K. R. K. and Wandell, B. A.	'Discrete Analysis of Spatial-Sensitivity Models'	*Journal of the Optical Society of America, Vol. 5, No. 5, 1988*	1893-1911
Olzak, L. A. and Thomas, J. P.	'Configural Effects Constraint Fourier Models of Pattern Discrimination'	*Vision Research, Vol. 32, No. 10, 1992*	
Olzak, L. A., Thomas, J. P., and Stanislaw, H.	'Development of a Chromatic/Luminance Contrast Scale'	*U. S. Department of Transportation, U. S. Coast Guard Office of Engineering and Development, 1987*	435-445
Olzak, L. A. and Thomas, J. P.	'What can discrimination tasks tell us about texture perception?	*OSA Annual Meeting Technical Digest, Vol. 23, 1992*	111-147

Olzak, L. A. and Thomas, J. P.	'When orthogonal orientations are not processed independently'	*Vision Research, Vol. 31, No. 1, 1991*	
Olzak, L. A., Wickens, T. D., and Thomas, J. P.	'Constraints on Fourier Models of Human Pattern Recognition'	*SPIE Vol. 1666 Human Vision, Visual Processing, and Digital Display III, 1992*	77-81
Overington, I.	'Towards a Complete Model of Photopic Visual Threshold Performance'	*Optical Engineering, Vol. 21, No. 1, January-February 1982*	
Page, Don H.	'Geometrical description of Berry's phase'	*Physical Review A, V36, No. 7, October 1, 1987*	786-804
Park, S. K.	'Image Gathering, Interpolation and Restoration: A Fidelity Analysis'	*SPIE, 1992 Technical Symposium on Visual Information Processing, 20-24 April 1992, Orlando*	
Petersen, H. E., D. J. Dugas	'The Relative Importance of Contrast and Motion in Visual Detection'	*Human Factors, 14(3), 1972*	1-103,
Pollen, D. A., J. P. Gaska, L. D. Jacobson	'Responses of Simple and Complex Cells to Compound Sine-Wave Gratings'	*Vision Research, Vol. 28, No. 1, 1988*	
Porat, M., and Y. Y. Zeevi	'The Generalized Gabor Scheme of Image Representation in Biological and Machine Vision'	*IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10, No. 4, July 1988*	
Portilla, J., R. Navarro, O. Nestares	'Texture synthesis-by-analysis method based on a multi-scale early-vision model'	*Opt. Eng. 35(8), August 1996*	
Provost, J. P., and G. Vallee	'Riemannian Structure on Manifolds of Quantum States'	*Communications in Mathematical Physics, Vol. 76, 1980*	
Quick R. F., Jr.	'A Vector-Magnitude Model of Contrast Detection'	*Kybernetic, 1974, 16*	
Reed, T. R.	'Segmentation of Textured Images and Gestalt Organization Using Spatial/Spatial-Frequency Representations'	*IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, January 1990*	29-65

Reid, R. C., R. Shapley	'Brightness Induction by Local Contrast and the Spatial Dependence of Assimilation'	*Vision Research, Vol. 28, No. 1, 1988*	
Rolland, J. P., H. H. Barrett	'Effect of random background inhomogeneity on observer detection performance'	*J. Opt. Soc. Am. A., Vol. 9, No. 5, May 1992*	141-166
Robinson, G. H.	'Visual Search by Automobile Drivers'	*Human Factors, 14(4), 1972*	249-273
Rotman, S. R., E. SW. Gordon, M. L. Kowalczyk	'Modeling Human Search and Target Acquisition Performance: I. First Detection Probability in a Realistic Multi-target'	*Optical Engineering, 28(11), November 1989*	
Rotman, S. R., E. S. Gordon, M. L. Kowalcyzk	'Modeling Human Search and Target Acquisition Performance: III. Target Detection in the Presence of Obscurants'	*Optical Engineering, June 1991*	677-688
Rotman, S. R., E. S. Gordon, O. Hadar, N. S. Kopeika, V. George, M. L. Kowalczyk	'Modeling Human Search and Target Acquisition Performance: V. Search Strategy'	*Optical Engineering, February 1993*	
Ruppeiner, G.	'Thermodynamics: A Riemannian geometric model'	*Physical Review, Vol. 20, No. 4, October 1979*	89-101
Sadot, D., N. S. Kopeika, S. R. Rotman	'Incorporation of Atmospheric Blurring Effects in Target Acquisition Modeling of Thermal Images'	*Infrared Phys. Technol. Vol. 36, No. 2, 1995*	1895-1900
Saito, N., G. Beyikin	'Multi-resolution Representations using the Auto-Correlation Functions of Compactly Supported Wavelets'	*IEEE Transactions on Signal Processing, 15 August 1991*	1977-1990
Schade, O. H., Sr.	'Optical and Photoelectric Analog of the Eye'	*Journal of the Optical Society of America, Vol. 46, No. 9, September 1956*	
Schmieder, D. E. and Weahersby, M. R.	'Detection performance in clutter with variable resolution'	*IEEI Transactions on Aerospace and Electronic Systems, Vol. AES-19, No. 4, 1983*	
Schnitzler, A. D.	'Image-detector model and parameters of the human visual	*Journal of the Optical Society of America, Vol.*	860-865

	system'	*63, No. 11, November 1973*	
Schwartz, E. L.	'Computational Anatomy and Functional Architecture of Striate Cortex: A Spiral Mapping Approach to Perceptual Coding'	*Vision Research, Vol. 20, 1980, Printed in Great Britain*	811-820
Simonotto, E., M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss	'Visual Perception of Stochastic Resonance'	*Physical Review Letters, Vol. 78, No. 6, February 10, 1997*	1533-1542
Swets, J. A.	'Decision Processes in Perception'	*Psychological Review, Vol. 68, No. 5, 1961*	513-518
Swets, J. A.	'The Relative Operating Characteristic in Psychology'	*Science, Vol. 182, December 1973*	
Swindale, N. W.	'The development of topography in a visual cortex: a review of models'	*Network: Computation in Neural Systems, 7(1996), Printed in the UK*	
Tanner, W. P., Jr.	'A Decision-Making Theory of Visual Detection'	*Psychological Review, Vol. 61, No. 6, 1954*	
Thomas, J. P.	'Underlying psychometric functions for detecting gratings and identifying spatial frequency'	*J. Opt. Soc. Am., Vol. 73, No. 6, June 1983*	
Thomas, J. P., J. Gille, R. A. Barker	'Simultaneous visual detection and identification: theory and data'	*J. Opt. Soc. Am., Vol. 72, No. 12, December 1982*	1-58
Thomas, J. P., L. Kerr	'Effect of ramp-like contours upon perceived size and detection threshold'	*Perception & Psychophysics, Vol. 5(6), 1969*	1-13
Thuang, J., C. Beckman, M. Abrahamsson, J. Sjöstrand	'The 'Light Scattering Factor' '	*Investigative Ophthalmology & Visual Science, Vol. 36, No. 11, October 1995*	1-20
Tootell, R. BH., S. L. Hamilton, M. S. Silverman, E. Switkes	'Functional Anatomy of Macaque Striate Cortex. I. Ocular Dominance, Binocular Interactions, and Baseline Conditions'	*The Journal of Neuroscience, 8(5), May 1988*	193-200
Tootell, R. G. H., E. Switkes, M. S. Silverman, and S.	'Functional Anatomy of Macaque Striate Cortex. II. Retinotopic Organization'	*The Journal of Neuroscience, 8(5), May 1988*	1-21

L. Hamilton			
Tootell, R. G. H., E. Switkes, M. S. Silverman, S. L. Hamilton, R. L. De Valois, and E. Switkes	'Functional Anatomy of Macaque Striate Cortex. III. Color'	*The Journal of Neuroscience, 8(5), May 1988*	1-4
Tootell, R. B. H., S. L. Hamilton, and E. Switkes	'Functional Anatomy of Macaque Striate Cortex. IV. Contrast and Magno-Parvo Streams'	*The Journal of Neuroscience, 8(5), May 1995*	1-9
Tootell, R. G. H., M. S. Silverman, S. L. Hamilton, E. Switkes, and Russell L. De Valois	'Functional Anatomy of Macaque Striate Cortex. V. Spatial Frequency'	*The Journal of Neuroscience, 8(5), May 1988*	130-139
Treisman, Anne	'Preattentive Processing Vision'	*Computer Vision, Graphics, and Image Processing 31, 1985*	722-726
Proposal -	'Center for Virtual Proving Ground Simulation: Mechanical and Electromechanical Systems'	*University of Iowa and University of Texas-Austin, 28 October 1996*	320-334
Udin, S. B.	'Formation of Topographic Maps'	*Ann. Rev. Neurosci., 11, 1988*	
Van Essen, D. C. and Anderson, C. H.	'Information processing strategies and pathways in the primate retina and visual cortex'	*An Introduction to Neural and Electronic Networks, 1990*	
Vos, J. J.	'On the relation between various levels of target acquisition'	*Netherlands Organization for Applied Scientific Research, 1989-38*	
Vos, J. J., and Van Meeteren, A.	'Phind, an analytical model to predict target acquisition distance with image intensifiers'	*Netherlands Organization for Applied Scientific Research, 1989-45*	545-553
Waldman, G., Wootton, J., Hobson, G., and Luetkemeyer, K.	'A normalized clutter measure for images'	*Computer Vision, Graphics, and Image Processing, Vol. 42, 1988*	
Waldman, G., Wootton, J., and Hobson, G.	'Visual detection with search: an empirical model'	*IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21,*	

		No. 3, 1991	
Watson, A. B.	'Probability Summation Over Time'	*Vision Research, Vol. 19, 1979*	1855-1862
Weldon, T. P., W. E., Higgins, and D. F. Dunn	'Gabor filter design for multiple texture segmentation'	*Optical Engineering, Vol. 35, No. 10, October 1996*	
Williams, D.	'Progress in Vision Research'	*Optics & Photoptics News, August 1991*	743-755
Wilson, H. R. and Richard, W. A.	'Curvature and separation discrimination at texture boundaries'	*Optical Society of America, Vol. 9, No. 10, October 1992*	
Zhu, Y. M., and R. Goutte	'Analysis and comparison of space/spatial-frequency and multi-scale methods for texture segmentation'	*Optical Engineering, Vol. 34, No. 1, January 1995*	416-423